

Alexander S. Holevo

Quantum Systems, Channels, Information

A Mathematical Introduction

De Gruyter

Physics and Astronomy Classification Scheme 2010: 03.67.-a; 05.30.-d; 02.30.Tb; 02.50.-r

ISBN 978-3-11-027325-0
e-ISBN 978-3-11-027340-3

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

© 2012 Walter de Gruyter GmbH, Berlin/Boston

Typesetting: Da-TeX Gerd Blumenstein, Leipzig, www.da-tex.de

Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen

∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com



Preface

Quantum information theory studies the general laws of transfer, storage and processing information in systems obeying the laws of quantum mechanics. It took shape as a self-consistent area of research in the 1990s, while its origin can be traced back to the 1950–1960s, which was when the basic ideas of reliable data transmission and of Shannon’s information theory were developed. At the first stage, which covers the period 1960–1980, the main issue consisted of the fundamental restrictions on the possibilities of information transfer and processing caused by the quantum-mechanical nature of its carrier. Modern technological developments, relying upon the achievements of quantum electronics and quantum optics, suggest that in the foreseeable future such restrictions will become the main obstacle limiting further extrapolation of existing technologies and principles of information processing.

The emergence, in the 1980–1990s, of the ideas of the quantum computing, quantum cryptography and the new communication protocols, on the other hand, allowed discussing not only the restrictions, but also the new possibilities created by the use of specific quantum resources, such as quantum entanglement, quantum complementarity, and quantum parallelism. Quantum information theory provides the clue to understanding these fundamental issues, and stimulates the development of experimental physics, with potential importance to new, effective applications. At present, investigations in the area of *quantum information science*, including information theory, its experimental aspects and technological developments, are ongoing in advanced research centers throughout the world.

The mathematical toolbox of “classical” information theory contains methods based in probability theory, combinatorics, modern algebra, including algebraic geometry, etc. For a mathematician sensible to the impact of his research on the natural sciences, information theory can be a source of deep ideas and new, challenging problems, with sound motivation and applications. This equally, if not to a greater extent, applies to quantum information theory, the scope of which turns out to be closely connected to multilinear algebra and non-commutative analysis, convexity and asymptotic theory of finite-dimensional normed spaces, subtle aspects of positivity and tensor products in operator algebras, and with the methods of random matrices. Nowadays, the intimate connections to operator spaces and so called “quantum functional analysis” have been revealed and explored.

In 2002, the Moscow Independent University published the author’s lecture notes (in Russian), in which an attempt was made at a mathematician’s introduction to problems of quantum information theory. In 2010, a substantially expanded text was pub-

lished with the title “Quantum systems, channels, information”. The author’s intention was to provide a widely accessible and self-contained introduction to the subject, starting from primary structures and leading up to nontrivial results with rather detailed proofs, as well as to some open problems. The present English text is a further step in that direction, extending and improving on the Russian version of 2010.

The exposition is organized in concentric circles, the N -th round consisting of Parts I to N , where each circle is self-contained. The reader can restrict himself to any of these circles, depending on the depth of presentation that he demands. In particular, in Part I to Part IV, we consider finite-dimensional systems and channels, whereas the infinite dimensional case is treated in the final Part V.

Part I starts with a description of the statistical structure of quantum theory. After introducing the necessary mathematical prerequisites in Chapter 1, the central focus in Chapters 2, 3 is on the discussing the key features of the quantum *complementarity* and *entanglement*. The former is reflected by the noncommutativity of the algebra of observables of the system, while the latter is reflected by the tensor product structure of composite quantum systems. Chapter 3 also contains first applications of the information-theoretic approach to quantum systems.

In information theory, the notions of a *channel* and its *capacity*, giving a measure of ultimate information-processing performance of the channel, play a central role. In Chapter 4 of Part II, a review of the basic concepts and necessary results from classical information theory is provided, the quantum analogs of which are the main subject of the following chapters. The concepts of *random coding* and *typicality* are introduced, and then extended to the quantum case in Chapter 5. That chapter contains direct and self-consistent proofs of the quantum information bound and of the primary coding theorems for the classical-quantum channels, which will later serve as a basis for the more advanced capacity results in Chapter 8.

Part III is devoted to the study of quantum channels and their entropy characteristics. In Chapter 6, we discuss the general concept and structure of a quantum channel, with the help of a variety of examples. From the point of view of operator algebras, these are normalized, completely positive maps, the analog of Markov maps in non-commutative probability theory, and play the role of morphisms in the category of quantum systems. From the point of view of statistical mechanics, a channel gives an overall description of the evolution of an open quantum system interacting with an environment – a physical counterpart of the mathematical dilation theorem. Various entropic quantities essential to characterize the information-processing performance, as well as the irreversibility of the channel, are investigated in Chapter 7.

Part IV is devoted to the proofs of advanced coding theorems, which give the main capacities of a quantum channel. Remarkably, in the quantum case, the notion of the channel capacity splits, giving a whole spectrum of information-processing characteristics, depending on the kind of data transmitted (classical or quantum), as well as on the additional communication resources. In Chapter 8, we discuss the *classical capacity* of a quantum channel, i.e. the capacity for transmitting classical data. We

touch upon the tremendous progress made recently in the solution of the related *additivity problem* and point out the remaining questions. Chapter 9 is devoted to the classical entanglement-assisted capacity and its comparison with unassisted capacity. In Chapter 10, we consider reliable transmission of quantum information (i.e. quantum states), which turns out to be closely related to the private transmission of classical information. The corresponding coding theorems provide the *quantum capacity* and the *private classical capacity* of a quantum channel.

In Part V, we pass from finite-dimensional to separable Hilbert space. Chapter 11 deals with the new obstacles characteristic for infinite-dimensional channels – singular behavior of the entropy (infinite values, discontinuity), and the emergence of the input channel constraints (e.g. finiteness of the signal energy) and of the continuous optimizing state ensembles. Chapter 12 treats the bosonic Gaussian systems and channels on the canonical commutation relations (many experimental demonstrations of quantum information processing were realized in such “continuous variables” systems, based in particular on the principles of quantum optics). We assume the reader has some minor background in the field and start with a rather extended introduction at the beginning of Chapter 12. Next, we describe and study in detail the Gaussian states and channels. The main mathematical problems here are the structure of the multi-mode quantum Gaussian channels and the computation of the various entropic quantities characterizing their performance. While the classical entanglement-assisted capacity is, in principle, computable for a general Gaussian channel, the quantum capacity is found only for restricted classes of channels, and the unassisted classical capacity in general presents an open analytical problem, namely that of verifying the conjecture of “quantum Gaussian optimizers”, which is comparable in complexity to the additivity problem (also open for the class of Gaussian channels), and appears to be closely related to it.

This book does not intend to be an all-embracing text in quantum information theory and its content definitely reflects the author’s personal research interests and preferences. For example, the important topics of entanglement quantification and error correction are mentioned only briefly. An interested reader can find an account of these in other sources, listed in the notes and references to the individual chapters. Quantum information theory is in a stage of fast development and new, important results continue to appear. Yet, we hope the present text will be a useful addition to the existing literature, particularly for mathematically inclined readers eager to penetrate the fascinating world of quantum information.

The basis for these lecture notes was a course taught by the author at the Moscow Institute of Physics and Technology, Moscow State University, and several Western institutions. The author acknowledges stimulating discussions, collaborations and invaluable support of R. Ahlswede, A. Barchielli, C. H. Bennett, G. M. D’Ariano, C. Fuchs, V. Giovannetti, O. Hirota, R. Jozsa, L. Lanz, O. Melsheimer, H. Neumann, M. B. Ruskai, P. W. Shor, Yu. M. Suhov, K. A. Valiev, R. Werner, A. Winter, M. Wolf.

I extend special thanks to my colleagues Maxim Shirokov and Andrey Bulinsky for their careful reading of the manuscript and the suggestion of numerous improvements.

This work was supported by Russian Foundation for Basic Research, Fundamental Research Programs of the Russian Academy of Sciences, and by the Cariplo Fellowship organized by the Landau Network – Centro Volta.

Contents

Preface	v
I Basic structures	
1 Vectors and operators	3
1.1 Hilbert space	3
1.2 Operators	4
1.3 Positivity	5
1.4 Trace and duality	6
1.5 Convexity	8
1.6 Notes and references	9
2 States, observables, statistics	10
2.1 Structure of statistical theories	10
2.1.1 Classical systems	10
2.1.2 Axioms of statistical description	11
2.2 Quantum states	14
2.3 Quantum observables	16
2.3.1 Quantum observables from the axioms	16
2.3.2 Compatibility and complementarity	18
2.3.3 The uncertainty relation	21
2.3.4 Convex structure of observables	22
2.4 Statistical discrimination between quantum states	25
2.4.1 Formulation of the problem	25
2.4.2 Optimal observables	25
2.5 Notes and references	31
3 Composite systems and entanglement	34
3.1 Composite systems	34
3.1.1 Tensor products	34
3.1.2 Naimark's dilation	36
3.1.3 Schmidt decomposition and purification	38

3.2	Quantum entanglement vs “local realism”	41
3.2.1	Paradox of Einstein–Podolski–Rosen and Bell’s inequalities	41
3.2.2	Mermin–Peres game	45
3.3	Quantum systems as information carriers	47
3.3.1	Transmission of classical information	47
3.3.2	Entanglement and local operations	48
3.3.3	Superdense coding	49
3.3.4	Quantum teleportation	50
3.4	Notes and references	52

II The primary coding theorems

4	Classical entropy and information	57
4.1	Entropy of a random variable and data compression	57
4.2	Conditional entropy and the Shannon information	59
4.3	The Shannon capacity of the classical noisy channel	62
4.4	The channel coding theorem	64
4.5	Wiretap channel	69
4.6	Gaussian channel	71
4.7	Notes and references	72
5	The classical-quantum channel	74
5.1	Codes and achievable rates	74
5.2	Formulation of the coding theorem	75
5.3	The upper bound	78
5.4	Proof of the weak converse	83
5.5	Typical projectors	87
5.6	Proof of the Direct Coding Theorem	92
5.7	The reliability function for pure-state channel	95
5.8	Notes and references	98

III Channels and entropies

6	Quantum evolutions and channels	103
6.1	Quantum evolutions	103
6.2	Completely positive maps	106
6.3	Definition of the channel	112

6.4	Entanglement-breaking and PPT channels	114
6.5	Quantum measurement processes	117
6.6	Complementary channels	119
6.7	Covariant channels	124
6.8	Qubit channels	127
6.9	Notes and references	129
7	Quantum entropy and information quantities	132
7.1	Quantum relative entropy	132
7.2	Monotonicity of the relative entropy	133
7.3	Strong subadditivity of the quantum entropy	138
7.4	Continuity properties	140
7.5	Information correlation, entanglement of formation and conditional entropy	142
7.6	Entropy exchange	147
7.7	Quantum mutual information	149
7.8	Notes and references	151
IV	Basic channel capacities	
8	The classical capacity of quantum channel	155
8.1	The coding theorem	155
8.2	The χ -capacity	157
8.3	The additivity problem	160
8.3.1	The effect of entanglement in encoding and decoding	160
8.3.2	A hierarchy of additivity properties	164
8.3.3	Some entropy inequalities	166
8.3.4	Additivity for complementary channels	169
8.3.5	Nonadditivity of quantum entropy quantities	171
8.4	Notes and references	178
9	Entanglement-assisted classical communication	180
9.1	The gain of entanglement assistance	180
9.2	The classical capacities of quantum observables	184
9.3	Proof of the Converse Coding Theorem	188
9.4	Proof of the Direct Coding Theorem	190
9.5	Notes and references	194

10	Transmission of quantum information	195
10.1	Quantum error-correcting codes	195
10.1.1	Error correction by repetition	195
10.1.2	General formulation	197
10.1.3	Necessary and sufficient conditions for error correction	198
10.1.4	Coherent information and perfect error correction	200
10.2	Fidelities for quantum information	203
10.2.1	Fidelities for pure states	203
10.2.2	Relations between the fidelity measures	205
10.2.3	Fidelity and the Bures distance	208
10.3	The quantum capacity	210
10.3.1	Achievable rates	210
10.3.2	The quantum capacity and the coherent information	215
10.3.3	Degradable channels	217
10.4	The private classical capacity and the quantum capacity	220
10.4.1	The quantum wiretap channel	220
10.4.2	Proof of the Private Capacity Theorem	223
10.4.3	Large deviations for random operators	229
10.4.4	The Direct Coding Theorem for the quantum capacity	232
10.5	Notes and references	237
V	Infinite systems	
11	Channels with constrained inputs	243
11.1	Convergence of density operators	243
11.2	Quantum entropy and relative entropy	247
11.3	Constrained c-q channel	249
11.4	Classical-quantum channel with continuous alphabet	252
11.5	Constrained quantum channel	254
11.6	Entanglement-assisted capacity of constrained channels	257
11.7	Entanglement-breaking channels in infinite dimensions	259
11.8	Notes and references	264
12	Gaussian systems	266
12.1	Preliminary material	266
12.1.1	Spectral decomposition and Stone's Theorem	266
12.1.2	Operators associated with the Heisenberg commutation relation	269

12.1.3	Classical signal plus quantum noise	272
12.1.4	The classical-quantum Gaussian channel	275
12.2	Canonical commutation relations	276
12.2.1	Weyl–Segal CCR	276
12.2.2	The symplectic space	279
12.2.3	Dynamics, quadratic operators and gauge transformations	281
12.3	Gaussian states	284
12.3.1	Characteristic function	284
12.3.2	Definition and properties of Gaussian states	285
12.3.3	The density operator of Gaussian state	289
12.3.4	Entropy of a Gaussian state	290
12.3.5	Separability and purification	293
12.4	Gaussian channels	296
12.4.1	Open bosonic systems	296
12.4.2	Gaussian channels: basic properties	300
12.4.3	Gaussian observables	301
12.4.4	Gaussian entanglement-breaking channels	303
12.5	The capacities of Gaussian channels	307
12.5.1	Maximization of the mutual information	307
12.5.2	Gauge-covariant channels	308
12.5.3	Maximization of the coherent information	310
12.5.4	The classical capacity: conjectures	311
12.6	The case of one mode	314
12.6.1	Classification of Gaussian channels	314
12.6.2	Entanglement-breaking channels	320
12.6.3	Attenuation/amplification/classical noise channel	321
12.6.4	Estimating the quantum capacity	325
12.7	Notes and references	329
	Bibliography	333
	Index	346

Part I

Basic structures

Chapter 1

Vectors and operators

We will deal with quantum-mechanical systems described by finite-dimensional Hilbert spaces. On the one hand, one can see already in that case, and perhaps most manifestly, the distinctive features of quantum statistics. On the other hand, finite-level systems are of the main interest to applications such as quantum communication, cryptography, and computing (although in the field of quantum information theory attention was recently attracted by the “continuous-variables systems,” that are described by infinite-dimensional spaces, see Chapters 11, 12).

1.1 Hilbert space

Let \mathcal{H} be a d -dimensional complex linear space, $\dim \mathcal{H} = d < \infty$, with the inner product $\langle \phi | \psi \rangle$, $\phi, \psi \in \mathcal{H}$, satisfying the axioms of unitary space (see e.g. [138]). However, following physical rather than mathematical tradition, we assume $\langle \phi | \psi \rangle$ to be linear with respect to the second argument ψ and antilinear with respect to ϕ . Again following a tradition in physics literature, we shall call it a Hilbert space (although emphasis of the mathematical Hilbert space theory is on the infinite-dimensional case).

We shall use Dirac’s notations: a vector $\psi \in \mathcal{H}$ (which conventionally should be thought of as a column vector) will often be denoted as $|\psi\rangle$. Correspondingly, $\langle \phi|$ will denote the linear function on \mathcal{H} , defined by the inner product

$$\langle \phi| : \psi \rightarrow \langle \phi | \psi \rangle, \quad \psi \in \mathcal{H}.$$

These linear functions themselves form another Hilbert space – the dual space \mathcal{H}^* of \mathcal{H} (and they should be thought of as row vectors $\langle \phi| = |\phi\rangle^*$, the Hermitian conjugate (adjoint) to $|\phi\rangle$). The space \mathcal{H}^* is anti-isomorphic to \mathcal{H} via the correspondence $|\psi\rangle \leftrightarrow \langle \psi|$. Then the inner product $\langle \phi | \psi \rangle$ is naturally thought of as the product of the “bra” and “ket” vectors $\langle \phi|$ and $|\psi\rangle$. The square of the norm, both in \mathcal{H} and \mathcal{H}^* , is $\langle \psi | \psi \rangle = \|\psi\|^2$.

This notation also allows for a convenient description of operators. For example, the “outer product” of “ket” and “bra” vectors $A = |\psi\rangle\langle \phi|$ describes the rank one operator, which acts onto vector $|\chi\rangle$ according to the formula $A|\chi\rangle = |\psi\rangle\langle \phi|\chi\rangle$. Let $\{e_i\}_{i=1,d}$ be an orthonormal basis in \mathcal{H} . The decomposition of arbitrary vector $\psi \in \mathcal{H}$ can then be written as

$$|\psi\rangle = \sum_{i=1}^d |e_i\rangle \langle e_i | \psi \rangle, \quad (1.1)$$

which is the same as

$$\sum_{i=1}^d |e_i\rangle\langle e_i| = I, \quad (1.2)$$

where I is the unit operator in \mathcal{H} .

Exercise 1.1. Write the matrix representation for operators in \mathcal{H} , similar to the representation (1.1) for vectors.

An additional advantage of Dirac's notations is that we do not need to write the symbol of the vector, leaving only the label(s), e.g. we may write $|i\rangle$ instead of $|e_i\rangle$, etc.

Sometimes we shall consider a *real* Hilbert (i.e. Euclidean) space. A fundamental difference with the complex case is revealed by the *polarization identity*

$$\beta(\phi, \psi) = \frac{1}{4} \sum_{k=0}^3 (-i)^k \beta(\phi + i^k \psi, \phi + i^k \psi), \quad (1.3)$$

which allows us to uniquely restore all values of the form $\beta(\phi, \psi)$, which is linear with respect to the second argument and antilinear with respect to the first, from its diagonal values $\beta(\psi, \psi)$, $\psi \in \mathcal{H}$ (in the real case, such a restoration is possible only for symmetric forms). Due to this, in order to prove an operator equality $A = B$, it is sufficient to establish the equality for all diagonal values $\langle \psi | A \psi \rangle = \langle \psi | B \psi \rangle$, $\psi \in \mathcal{H}$.

1.2 Operators

If A is an operator in \mathcal{H} (throughout this book we deal only with *linear* operators), then A^* denotes its adjoint, defined by

$$\langle \phi | A^* \psi \rangle = \langle A \phi | \psi \rangle \quad \phi, \psi \in \mathcal{H}. \quad (1.4)$$

An operator A is called *Hermitian* if $A = A^*$. An (orthogonal) *projector* is a Hermitian operator P such that $P^2 = P$. The range of projector P is the subspace

$$\mathcal{L} = \{ \psi : P|\psi \rangle = |\psi \rangle \}.$$

If $\|\psi\| = 1$, then $|\psi\rangle\langle\psi|$ is the projector onto the unit vector $|\psi\rangle$. More generally, for an orthonormal system $\{e_i\}_{i \in I}$, the operator $\sum_{i \in I} |e_i\rangle\langle e_i| = P$ is the projector onto the subspace generated by $\{e_i\}_{i \in I}$.

A *unitary operator* is an operator U such that $U^*U = I$. In the finite-dimensional case, which we here consider, this implies $UU^* = I$. A *partial isometry* is an operator U such that $U^*U = P$ is a projector. In this case $UU^* = Q$ also is a projector. U maps the range of P onto the range of Q *isometrically*, i.e. preserving inner products and norms of vectors.

Theorem 1.2 (Spectral Decomposition). *For a Hermitian operator A , there is an orthonormal basis of eigenvectors to which correspond real eigenvalues a_i , such that*

$$A = \sum_{i=1}^d a_i |e_i\rangle\langle e_i|. \quad (1.5)$$

Another useful form of the spectral decomposition is obtained if we consider the *distinct* eigenvalues $\{a\}$ and the corresponding spectral projectors

$$E_a = \sum_{i:a_i=a} |e_i\rangle\langle e_i|.$$

The collection of distinct eigenvalues $\text{spec}(A) = \{a\}$ is called the *spectrum* of the operator A . Then

$$A = \sum_{a \in \text{spec}(A)} a E_a. \quad (1.6)$$

This representation is unique up to the ordering of the eigenvalues. The collection of projectors $\{E_a\}$ forms an *orthogonal resolution of the identity*:

$$E_a E_{a'} = \delta_{aa'} E_a, \quad \sum_{a \in \text{spec}(A)} E_a = I. \quad (1.7)$$

The Hilbert space \mathcal{H} splits into the direct orthogonal sum of the ranges of the projectors $\{E_a\}$, on which A acts as multiplication by a .

Both unitary and Hermitian operators are special cases of *normal* operators, satisfying $A^* A = A A^*$. For normal operators, the spectral decomposition (1.5) holds, with complex eigenvalues a_i .

1.3 Positivity

A Hermitian operator A is called *positive*, $A \geq 0$, if $\langle \psi | A \psi \rangle \geq 0$ for all $\psi \in \mathcal{H}$. The eigenvalues of a positive operator are all nonnegative: $a \geq 0$ for $a \in \text{spec}(A)$.

The operator is positive if and only if it can be represented as $A = B^* B$ for some operator B . A positive operator has a unique positive square root, i.e. for any positive operator A there exists one and only one positive operator B , denoted by $\sqrt{A} = A^{1/2}$ such that $B^2 = A$.

For an arbitrary Hermitian operator A , we have

$$A = A_+ - A_-, \quad (1.8)$$

where $A_+ = \sum_{a>0} a E_a$, $A_- = -\sum_{a<0} a E_a$ are positive operators called the *positive* and *negative* part of the operator A .

Theorem 1.3 (Polar Decomposition). *Any operator A in \mathcal{H} allows the decomposition*

$$A = U|A| = |A^*|U, \quad (1.9)$$

where $|A| = \sqrt{A^*A}$ is a positive operator and U is unitary.

The *support* $\text{supp } A$ of a positive operator A is the span of eigenvectors of A corresponding to positive eigenvalues. In the polar decomposition, the unitary operator is uniquely defined only on $\text{supp } A$.

Occasionally, we will make use of a *real* Hilbert space, i.e. Euclidean space, and the operators in it. In that case, the Hermitian operators are replaced by symmetric operators, and unitary operators by orthogonal ones, with formally similar definitions. The polar decomposition of an operator A holds, with $|A|$ symmetric positive and U an orthogonal operator.

1.4 Trace and duality

The *trace* of an operator T is defined by

$$\text{Tr } T = \sum_{i=1}^d \langle e_i | T e_i \rangle, \quad (1.10)$$

where $\{e_i\}$ is an arbitrary orthonormal basis.

Exercise 1.4. Show that the definition does not depend on the choice of basis, and that

$$\text{Tr } A^* = \overline{\text{Tr } A}, \quad \text{Tr } AB = \text{Tr } BA, \quad (1.11)$$

where bar denotes the complex conjugate. Show that

$$\text{Tr } |\psi\rangle\langle\varphi|A = \langle\varphi|A|\psi\rangle. \quad (1.12)$$

Show that for $A, B \geq 0$

$$\text{Tr } AB \geq 0 \quad (1.13)$$

with equality if and only if $AB = 0$.

The expression

$$\|T\|_1 = \text{Tr } |T| \quad (1.14)$$

defines the *trace norm* on the complex linear space of all operators in \mathcal{H} . Notice that for Hermitian T

$$\|T\|_1 = \sum_{i=1}^d |t_i|, \quad (1.15)$$

where t_i are the eigenvalues of T . Another important norm is the *operator norm*

$$\|A\| \equiv \|A\|_\infty = \max_{\psi: \|\psi\|=1} \|A\psi\|, \quad (1.16)$$

which for a Hermitian A is equal to

$$\|A\| = \max_{a \in \text{spec}(A)} |a|. \quad (1.17)$$

There is an important inequality

$$|\text{Tr } TA| \leq \|T\|_1 \|A\|, \quad (1.18)$$

where equality can be attained either for any fixed T or for any fixed A : for a fixed T take $A = U^*$, where U is the unitary from the polar decomposition $T = |T|U$. Thus,

$$\|T\|_1 = \max_U |\text{Tr } TU| = \max_{\|A\|=1} |\text{Tr } TA|. \quad (1.19)$$

Similarly, for given A the equality is attained for $T = |\psi\rangle\langle\psi|U^*$, where U is the unitary from the polar decomposition $A = U|A|$, and ψ is the normalized eigenvector of $|A|$, with the maximal eigenvalue. It follows that

$$\|A\| = \max_{\|T\|_1=1} |\text{Tr } TA|. \quad (1.20)$$

These facts underly an important duality relation. The space of all operators in \mathcal{H} , equipped with the trace norm, becomes the complex Banach space $\mathfrak{T}(\mathcal{H})$. The same space equipped with the operator norm (and the product operation) is the Banach algebra $\mathfrak{B}(\mathcal{H})$. These two Banach spaces are in mutual duality, which means that every linear function on $\mathfrak{T}(\mathcal{H})$ has the form $T \rightarrow \text{Tr } TA$ for some $A \in \mathfrak{B}(\mathcal{H})$, with the norm given by (1.20), and, conversely, every linear function on $\mathfrak{B}(\mathcal{H})$ has the form $A \rightarrow \text{Tr } TA$ for some $T \in \mathfrak{T}(\mathcal{H})$ with the norm (1.19).

The subscript h will be applied to spaces of operators to denote the corresponding real Banach spaces of Hermitian operators. Then, the real Banach spaces $\mathfrak{T}_h(\mathcal{H})$ and $\mathfrak{B}_h(\mathcal{H})$ are again in the mutual duality, provided by the bilinear form $T, A \rightarrow \text{Tr } TA$. It follows from (1.11) that this form is real-valued for Hermitian T, A .

In the finite-dimensional case the spaces $\mathfrak{T}(\mathcal{H})$ and $\mathfrak{B}(\mathcal{H})$ coincide with the space of all operators in \mathcal{H} and differ only by the norms. In the infinite-dimensional case, however, they are substantially different (just as the classical l^1 and l^∞ spaces over finite versus infinite sets).

Exercise 1.5. Show that the complex dimensionality of the space of all operators in \mathcal{H} is d^2 , while the real dimensionality of the subspace of all Hermitian operators in \mathcal{H} is again d^2 . If \mathcal{H} is a real, d -dimensional Hilbert space, the dimensionality of the space of all operators in \mathcal{H} is d^2 , while the dimensionality of the subspace of all symmetric operators in \mathcal{H} is $d(d+1)/2$.

1.5 Convexity

A subset \mathfrak{S} of a real linear space is *convex* if, for any finite collection of points $\{S_j\} \subset \mathfrak{S}$ and any probability distribution $\{p_j\}$, the *convex combination* $S = \sum_{j=1} p_j S_j$ is in \mathfrak{S} (it is sufficient to impose this requirement only on collections of two points, which means that if the set \mathfrak{S} contains two points, it must contain the whole segment connecting these points). In a convex set, a special role is reserved for *extreme points*, which cannot be represented as a nontrivial convex combination of other points. This is equivalent to the fact that $S = pS_1 + (1-p)S_2$, $0 < p < 1$, implies $S = S_1 = S_2$, which means that there is no segment in \mathfrak{S} containing S as its interior point. We denote by $\text{ext}(\mathfrak{S})$ the set of all extreme points of a convex set \mathfrak{S} . The following general result holds:

Theorem 1.6 (Caratheodory). *Let \mathfrak{S} be a compact convex subset of \mathbb{R}^n . In this case, every point $S \in \mathfrak{S}$ can be represented as a convex combination of at most $n + 1$ extreme points $S_j \in \text{ext}(\mathfrak{S})$:*

$$S = \sum_{j=1}^{n+1} p_j S_j, \quad S_j \in \text{ext}(\mathfrak{S}). \quad (1.21)$$

As an example consider the convex set \mathfrak{P}_n of all probability distributions $P = \{p_1, \dots, p_{n+1}\}$ on a set of $n + 1$ points. Its extreme points are the degenerate distributions, for which all p_j are zero, except for one which is equal to 1. There are $n + 1$ extreme points and every point of \mathfrak{P}_n is uniquely represented as their convex combination, with the coefficients p_j . This set is called *simplex*, and uniqueness of the representation is the characteristic feature of this convex set.

A real function \mathcal{F} on a convex subset \mathfrak{S} of a finite-dimensional linear space is *convex (concave)* if

$$\mathcal{F}\left(\sum_j p_j S_j\right) \leq (\geq) \sum_j p_j \mathcal{F}(S_j),$$

for an arbitrary convex combination of points $S_j \in \mathfrak{S}$. It is *affine* if it is both convex and concave. More generally, a map \mathcal{F} from one convex set into another is called *affine* if it preserves convex combinations:

$$\mathcal{F}\left(\sum_j p_j S_j\right) = \sum_j p_j \mathcal{F}(S_j).$$

Exercise 1.7. A continuous convex (in particular, affine) function on a compact convex set \mathfrak{C} attains its maximum at an extreme point of this set.

1.6 Notes and references

The basics of the operator formalism of quantum mechanics were laid down in Dirac's classical treatise [53]. An excellent introduction to finite quantum systems is given in Feynman's lectures [59]. From the vast variety of textbooks on linear algebra, we mention the modern course of Kostrikin and Manin [138], which takes into account the needs of quantum theory, and the book of Glazman and Ljubich [68], which is aimed at the active study of noncommutative, finite-dimensional analysis through solving problems. A variety of topics in modern matrix analysis, including many matrix inequalities, is covered in the book of Bhatia [27]. The fundamentals of convex analysis are exposed in, e.g. the books of Magaril–Il'jaev and Tikhomirov [154] and Rockafellar [172].

Chapter 2

States, observables, statistics

2.1 Structure of statistical theories

2.1.1 Classical systems

A classical system is described in terms of its *phase space* Ω , the points ω of which label deterministic states of the system. For simplicity, we consider in what follows finite sets Ω . By classical (statistical) *state* we denote a probability distribution $P = \{p_\omega\}$ on Ω . The collection of all probability distributions is the simplex $\mathfrak{P}(\Omega)$, in which every point is uniquely represented as a convex combination of extreme points – *pure states*, given by probability distributions that are degenerated in the points $\omega \in \Omega$. The simplest nontrivial system is the *bit*, the system with two pure states 0, 1, for which the set of statistical states is isomorphic to the unit segment of the real line.

Any random variable is a function $X = \{x_\omega\}$ on the phase space Ω , defining a decomposition of the space Ω into non-intersecting subsets Ω_x , in which X takes on the values x . The indicators $E_x(\omega)$ of these subsets satisfy the conditions

$$E_x(\omega)E_y(\omega) = \delta_{x,y}E_x(\omega); \quad \sum_x E_x(\omega) \equiv 1.$$

Along with random variables, which will henceforth be called *sharp observables*, one can consider *unsharp* observables which take the values x with probabilities $M(x|\omega)$ (such as observables with random error or intentionally randomized). The collection of conditional probabilities $M = \{M(x|\omega)\}$ is characterized by the properties

$$M(x|\omega) \geq 0; \quad \sum_x M(x|\omega) = 1.$$

For a sharp observable the probabilities $M(x|\omega) = E_x(\omega)$ are equal to 0 or 1, i.e. $M(x|\omega)^2 = M(x|\omega)$.

The probability distribution for the observable M in the state P is given by the formula

$$\mu_P^M(x) = \sum_\omega p_\omega M(x|\omega). \quad (2.1)$$

For a smooth transition to quantum systems it is useful to introduce a matrix representation of the classical quantities. Consider the Hilbert space spanned by a fixed

orthonormal basis $\{|\omega\rangle; \omega \in \Omega\}$. Given any function f_ω defined on Ω , we consider the diagonal operator

$$f = \sum_{\omega} f_{\omega} |\omega\rangle \langle \omega|.$$

Then a classical state gives rise to the operator $P = \sum_{\omega} p_{\omega} |\omega\rangle \langle \omega|$, characterized by the properties

$$P \geq 0; \quad \text{Tr } P = 1. \quad (2.2)$$

A classical observable gives rise to the resolution of the identity, i.e. a collection of operators $M = \{M_x\}$, where

$$M_x = \sum_{\omega} M(x|\omega) |\omega\rangle \langle \omega|,$$

satisfying the conditions

$$M_x \geq 0; \quad \sum_x M_x = I. \quad (2.3)$$

For a sharp observable, the operators $M_x = E_x$ are pairwise orthogonal projectors.

The relation (2.1) for the probability distribution of an observable can be rewritten as

$$\mu_P^M(x) = \text{Tr } P M_x. \quad (2.4)$$

2.1.2 Axioms of statistical description

At the basis of the mathematical structure of quantum theory, as well as of any other statistical theory, lies the separation of a statistical experiment into the two stages of preparation and measurement, which underlies the duality between states and observables. The following set of axioms, which is applicable to any statistical theory, stresses the parallel between the statistical description of classical and quantum systems.

Axiom 1. Let there be a given set \mathfrak{S} , whose elements are called *states*, and a set \mathfrak{M} , whose elements are called (finitely-valued) *observables*. For arbitrary $M \in \mathfrak{M}$, there is a finite set \mathcal{X}^M of outcomes. For arbitrary $S \in \mathfrak{S}$ and $M \in \mathfrak{M}$ there is a probability distribution μ_S^M on \mathcal{X}^M , called the *probability distribution* of the observable M in the state S .

The state S is interpreted as a more or less detailed description of the preparation of a *statistical ensemble*, i.e. a sample of independent individual representatives of the system under consideration, and the observable M , as a quantity that can be measured by a definite apparatus for each representative in the given ensemble. Axiom 1 thus presupposes the *reproducibility* of the individual experiments and the *stability of frequencies* under independent repetitions.



Figure 2.1. A quantum state described by a density operator S characterizes the preparation of a system, whereas the statistics of a measurement is described by a probability measure $\mu_S^M(x)$, where x labels the possible outcomes.

The following axiom expresses the possibility of the *mixing* of statistical ensembles.

Axiom 2. For arbitrary $S_1, S_2 \in \mathfrak{S}$ and an arbitrary number $p, 0 < p < 1$, there exists an $S \in \mathfrak{S}$ such that $\mu_S^M = p\mu_{S_1}^M + (1-p)\mu_{S_2}^M$ for all $M \in \mathfrak{M}$. S is called the *mixture* of states S_1 and S_2 with weights $p, 1-p$.

The next axiom describes the possibility of processing the information obtained from a measurement of an observable. Let $M_1, M_2 \in \mathfrak{M}$, and let f be a function from \mathcal{X}^{M_1} to \mathcal{X}^{M_2} such that, for all $S \in \mathfrak{S}$,

$$\mu_S^{M_2}(y) = \sum_{x: f(x)=y} \mu_S^{M_1}(x)$$

here, observable M_2 is called the *coarse-graining* of the observable M_1 . In this case, we write $M_2 = f \circ M_1$.

Axiom 3. For an arbitrary observable $M_1 \in \mathfrak{M}$ and an arbitrary (necessarily, finitely-valued) function f on \mathcal{X}^{M_1} , there exists an observable $M_2 \in \mathfrak{M}$, such that $M_2 = f \circ M_1$.

A pair of non-empty sets $(\mathfrak{S}, \mathfrak{M})$ satisfying the axioms 1–3 will be called *statistical model*. The statistical model is said to be *separated* if

Axiom 4. From the fact that $\mu_{S_1}^M = \mu_{S_2}^M$ for all $M \in \mathfrak{M}$ it follows that $S_1 = S_2$, and from $\mu_S^{M_1} = \mu_S^{M_2}$ for all $S \in \mathfrak{S}$ it follows that $M_1 = M_2$.

For a separated model both the operation of mixing in \mathfrak{S} and the coarse-graining in \mathfrak{M} are uniquely defined. Thus, the set of states \mathfrak{S} obtains a convex structure, and the set of observables \mathfrak{M} has the structure of a partial order.

Observables M_1, \dots, M_m are called *compatible* if they are all coarse-grainings of some observable M , that is, $M_j = f_j \circ M$ for $j = 1, \dots, m$. The outcome of compatible observables can be obtained as a result of data post-processing in a single-measurement experiment. Statistical models in which all observables are compatible are, in fact, classical.

Proposition 2.1. *Let $(\mathfrak{S}, \mathfrak{M})$ be a separated statistical model in which there exists the maximal observable M^* , such that all other observables are obtained by coarse-grainings of this maximal observable. In this case, there exist a finite set Ω , a one-to-one affine map $S \rightarrow P_S$ of the convex set of states \mathfrak{S} into the set of all probability distributions on Ω , and a one-to-one map $M \rightarrow f_M$ of the set of observables \mathfrak{M} onto a subset of random variables on Ω , preserving the relation of coarse-graining, such that*

$$\mu_S^M(x) = \sum_{\omega: f_M(\omega)=x} P_S(\omega). \quad (2.5)$$

Proof. Let Ω be the set of outcomes of the maximal observable M^* , and let $P_S(\omega)$ be its probability distribution in a state S . According to the assumption of maximality, for any observable M there is a function f_M , such that the relation (2.5) holds. The fact that the maps $S \rightarrow P_S$ and $M \rightarrow f_M$ are one-to-one follows from the separatedness of the model. Checking the properties of the maps is left as an exercise to the reader. \square

Thus the outcome of this “maximal observable” constitute the *phase space* Ω of the system, and all states are represented by probability measures on Ω . The main classical model is the *Kolmogorov model*, in which the set of states is a collection of all probability distributions $\mathfrak{P}(\Omega)$, and observables are described by random variables on Ω . The above proposition means that a statistical model in which all the observables are compatible can be embedded within the Kolmogorov model.

Another important classical model, which we call the *Wald model*, differs from Kolmogorov’s in that it embraces “unsharp” or “randomized” classical observables (see Section 2.1.1. Randomization was consistently used by Wald in statistical decision theory, and by von Neumann in the game theory). To obtain a result similar to Proposition 2.1, one should introduce stochastic versions of coarse-graining and compatibility. Let $M_1, M_2 \in \mathfrak{M}$, and Π be a transition probability from \mathcal{X}^{M_1} to \mathcal{X}^{M_2} , such that for all $S \in \mathfrak{S}$

$$\mu_S^{M_2}(y) = \sum_x \Pi(y|x) \mu_S^{M_1}(x).$$

In this case, observable M_2 is called the *stochastic coarse-graining* of the observable M_1 . Observables M_1, \dots, M_m are *stochastically compatible*, if they are all stochastic coarse-grainings of one observable M . In the Wald model, all observables are stochastically compatible and any separated model with this property can be embedded within a Wald model.

In the statistical model of quantum mechanics, states and observables are described by operators in a Hilbert space. Many efforts were spent on obtaining a reasonable axiomatic scheme, leading to the Hilbert space formalism of quantum mechanics. However, this is not our goal here and we will proceed in a different way. We take for granted the description of quantum states as density operators in a Hilbert space and

derive from it the most general concept of quantum observable (including both “sharp” and “unsharp” observables), following from the axioms of the statistical model.

2.2 Quantum states

A state of a quantum-mechanical system represents a statistical ensemble of independent, identically prepared copies of the system.

Definition 2.2. A *quantum state* is described by a density operator, i.e. a Hermitian operator S in a Hilbert space \mathcal{H} of the system, satisfying the conditions

$$S \geq 0, \quad \text{Tr } S = 1.$$

Let $\mathfrak{S}(\mathcal{H})$ be the set of all density operators. It is a convex subset of the real linear space of all Hermitian operators in \mathcal{H} . A convex combination $S = \sum_j p_j S_j$ of density operators will describe the *mixture* of the corresponding statistical ensembles, obtained by taking ensembles prepared in the states S_j and mixing them, with weights p_j . In a quantum statistical ensemble there are two kinds of randomness. One is due to fluctuations in the classical parameters of the preparation procedure. The other one is the intrinsically quantum irreducible randomness, present in any quantum state. The following theorem characterizes states without classical randomness.

Theorem 2.3. *Extreme points of the set of quantum states $\mathfrak{S}(\mathcal{H})$, called pure states, are precisely one dimensional projectors.*

Proof. The spectral decomposition of the Hermitian operator S reads

$$S = \sum_{i=1}^d s_i |e_i\rangle\langle e_i|, \quad s_i \geq 0, \quad \sum s_i = 1, \quad (2.6)$$

where s_i are the eigenvalues and $|e_i\rangle$ are the eigenvectors of S . If S is extreme, this sum can contain only one nonzero term, implying that S is a one-dimensional projector. Conversely, let S be a one-dimensional projector, and

$$S = pS_1 + (1-p)S_2, \quad 0 < p < 1.$$

Taking the square, we obtain

$$0 = S - S^2 = pS_1(I - S_1) + (1-p)S_2(I - S_2) + p(1-p)(S_1 - S_2)^2. \quad (2.7)$$

This equation is the sum of three positive operators, each of which must therefore be equal to zero. But this implies $S_1 = S_2 = S$, i.e. S is extreme. \square

From the viewpoint of probability theory, one-dimensional projectors are the non-commutative analog of distributions that are degenerated at some phase space point, while the role of the uniform distribution is played by the *chaotic state*, with the density operator $S = \frac{1}{d} I$.

Exercise 2.4. Show that if $\dim \mathcal{H} = d$, $\mathfrak{S}(\mathcal{H})$ can be embedded into the real vector space of dimensionality $n = d^2 - 1$. If \mathcal{H} is Euclidean (real Hilbert) space, then $n = d(d + 1)/2 - 1$.

The spectral decomposition (2.6) shows that every quantum state is a mixture of no more than d pure states, where $d = \dim \mathcal{H}$. Thus, in the case of a set of quantum states, the Caratheodory Theorem overestimates the number of extreme points that are actually present in the mixture (1.21). On the other hand, this theorem gives the exact value for the simplex of probability distributions on a “phase space” $\Omega = \{1, \dots, n + 1\}$, representing the statistical states of a classical system. This suggests the notion of interpreting quantum theory as a probabilistic model with specific nonclassical constraints (a hidden-variable theory). For a single quantum system, such a viewpoint is possible, although up to now it did not prove to be productive compared to standard quantum mechanics. On the other hand, for composite systems this viewpoint leads to an inevitable contradiction with the physical principle of locality (separability) (see Section 3.2.1).

The simplest, and yet fundamental example is *qubit* – a two-level quantum system, $\dim \mathcal{H} = 2$. We use the canonical basis: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. It is convenient to introduce the basis in the real space of 2×2 Hermitian matrices, called *Pauli matrices*

$$I \equiv \sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

In particular, a density operator $S \in \mathfrak{S}(\mathcal{H})$ can be represented as

$$S = \frac{1}{2}(I + a_x \sigma_x + a_y \sigma_y + a_z \sigma_z) = \frac{1}{2} \begin{bmatrix} 1 + a_z & a_x - i a_y \\ a_x + i a_y & 1 - a_z \end{bmatrix}. \quad (2.8)$$

The condition $\det S \geq 0$ imposes the following constraint onto the *Stokes parameters* $\vec{a} = (a_x, a_y, a_z)$:

$$|\vec{a}|^2 \equiv a_x^2 + a_y^2 + a_z^2 \leq 1.$$

Thus, $\mathfrak{S}(\mathcal{H})$ as a convex set is isomorphic to the unit ball in \mathbb{R}^3 , which we shall call the *Bloch ball*.

Exercise 2.5. Show that the density operator (2.8) has the eigenvalues $\frac{1 \pm |\vec{a}|}{2}$. Hint: use the results of Exercise 1.5.

The pure states are characterized by the condition $a_x^2 + a_y^2 + a_z^2 = 1$ and form the *Bloch sphere*. By introducing Euler angles θ, ϕ , so that $a_z = \cos \theta, a_x + ia_y = \sin \theta e^{i\phi}$, we have $S = |\psi(\vec{a})\rangle\langle\psi(\vec{a})|$, where

$$|\psi(\vec{a})\rangle = \begin{bmatrix} \cos \frac{\theta}{2} e^{-i\phi/2} \\ \sin \frac{\theta}{2} e^{i\phi/2} \end{bmatrix}. \quad (2.9)$$

For an electron, which is a spin-1/2 particle, the vector $\psi(\vec{a})$ describes the pure state, with the spin direction \vec{a} produced by a Stern-Gerlach filter with this direction of the magnetic field gradient. The mixed state, with $a_x = a_y = a_z = 0$, described by the density operator $S = \frac{1}{2}I$ for which all spin directions are equiprobable, is the chaotic state (see the Feynman lectures [59] for more detail).

Another important example of a two-level system is the polarization of a monochromatic photon, as visualized in classical optics experiments. In this case, the parameter $\frac{\theta}{2}$ is the angle of linear polarization, while $\frac{\phi}{2}$ characterizes circular polarization. For a linearly polarized photon $\phi = 0$. In particular, for a vertically polarized photon $\theta = 0$, while for a horizontally polarized one $\theta = \pi$.

2.3 Quantum observables

2.3.1 Quantum observables from the axioms

Consider a statistical model in which the state space is the convex set $\mathfrak{S}(\mathcal{H})$ of all density operators in a Hilbert space \mathcal{H} . Let M be an observable. Now, by the Axiom 2, the probabilities of the outcome of the measurement $\mu_S^M(x)$ should be *affine* functions of state.

Theorem 2.6. *Let the map $S \rightarrow \mu_S$ be an affine function on $\mathfrak{S}(\mathcal{H})$ such that $0 \leq \mu_S \leq 1$. In this case, there exists an operator M on \mathcal{H} with $0 \leq M \leq I$ such that, for every $S \in \mathfrak{S}(\mathcal{H})$*

$$\mu_S = \text{Tr } SM. \quad (2.10)$$

Proof (Sketch). Let us show that it is possible to extend the map $S \rightarrow \mu_S$ to a well-defined linear function on the space $\mathfrak{T}_h(\mathcal{H})$ of all Hermitian operators in \mathcal{H} .

First, we note that by separating positive and negative parts in the spectral decomposition (1.6), every Hermitian operator A can be represented as the difference of two positive operators, e.g. as in (1.8). Normalizing by the traces, we can always write

$$A = t_+ S_+ - t_- S_-, \quad t_{\pm} \geq 0, \quad S_{\pm} \in \mathfrak{S}(\mathcal{H}). \quad (2.11)$$

The representation of an operator A in the form of (2.11) is, of course, not unique.

Exercise 2.7. Using the affinity of the function $S \rightarrow \mu_S$, show that the value $\mu_A = t_+ \mu_{S_+} - t_- \mu_{S_-}$ is uniquely defined (i.e. depends solely on the operator A , but not on the concrete decomposition of A into the difference of positive operators), and that $A \rightarrow \mu_A$ is a real, linear function.

Next, every operator A can be represented as $A = A_1 + iA_2$, where $A_1 = \frac{1}{2}(A + A^*)$, $A_2 = \frac{1}{2i}(A - A^*)$ are Hermitian operators. This representation is unique, and by setting $\mu_A = \mu_{A_1} + i\mu_{A_2}$, we obtain the unique complex linear extension of the function μ_A to the space of all linear operators in \mathcal{H} . Any such function apparently has the form

$$\mu_A = \text{Tr } AM, \quad (2.12)$$

where M is an operator in H . Taking $A = |\psi\rangle\langle\psi|$, where ψ is a unit vector and using (1.12), we have, by assumption, $0 \leq \langle\psi|M\psi\rangle \leq 1$ for all ψ , hence $0 \leq M \leq I$. \square

A collection of operators $\{M_x, x \in \mathcal{X}\}$ is called the *resolution of the identity* or *probability operator-valued measure*¹ (POVM) on \mathcal{X} if

$$M_x^* = M_x \geq 0, \quad \sum_x M_x = I.$$

Corollary 2.8. Let $S \rightarrow \{\mu_S(x), x \in \mathcal{X}\}$ be an affine map of the convex set of quantum states $\mathfrak{S}(\mathcal{H})$ into the set of probability distributions on a finite set \mathcal{X} . In this case, there exists a POVM $M = \{M_x\}$ in \mathcal{H} such that

$$\mu_S(x) = \text{Tr } SM_x, \quad x \in \mathcal{X}. \quad (2.13)$$

The proof of the corollary is left as an *Exercise* to the reader. This corollary leads to the following

Definition 2.9. A quantum observable with outcomes in \mathcal{X} is described by a POVM $\{M_x, x \in \mathcal{X}\}$. The probability distribution μ_S^M of the observable $M = \{M_x\}$ in a state S is given by the formula (2.13).

By accepting this definition of a quantum observable, we introduce the *maximal* statistical model, with the state space coinciding with the set of all density operators in \mathcal{H} . In the standard expositions of quantum mechanics, the more restricted notion of quantum observable is used.

Definition 2.10. The quantum observable is called *sharp* if all operators $M_x = E_x$ are projectors: $E_x^2 = E_x$.

¹ The term “positive operator-valued measure” is also used in the literature, which is however less precise, as it does not reflect the normalization of the measure.

Exercise 2.11. Show that the condition $E_x^2 = E_x$, for all outcomes x , is equivalent to $E_x E_y = \delta_{xy} E_x$ for all x, y , i.e. the projectors E_x are mutually orthogonal.

The corresponding resolutions of the identity are called *orthogonal*. Sharp observables are thus described by orthogonal resolutions of the identity in \mathcal{H} . Observables for which the outcomes are real numbers, i.e. $\mathcal{X} \subset \mathbb{R}$, are called *real*. The spectral decomposition

$$X = \sum_{x \in \mathcal{X}} x E_x$$

sets a one-to-one correspondence between sharp real observables $E = \{E_x\}$ and Hermitian operators X in \mathcal{H} (which are also called observables). Coarse-graining for such observables takes the form of functional calculus. Namely, for a function $f : \mathbb{R} \rightarrow \mathbb{R}$, an observable $f \circ E$ corresponds to the Hermitian operator $f(X)$ (Exercise).

The *expectation* (mean value) of a sharp real observable X in the state S is given by the *Born–von Neumann statistical formula*

$$E_S(X) = \sum x \mu_S^E(x) = \text{Tr } SX.$$

The relation (2.13) can be regarded as an extension of this formula.

The statistical model in which states are described by density operators and observables – by Hermitian (self-adjoint) operators was considered in great detail by von Neumann [212].

2.3.2 Compatibility and complementarity

Definition 2.12. The *commutator* of the operators X, Y is $[X, Y] = XY - YX$. Operators X, Y commute if $[X, Y] = 0$.

Theorem 2.13. Let $E = \{E_x\}$ and $F = \{F_y\}$ be sharp observables. In this case, the following are equivalent:

- i. E, F are compatible;
- ii. projectors E_x and F_y commute for all x, y ;

If E, F are real observables, this is equivalent to

- iii. the corresponding Hermitian operators $X = \sum_x x E_x, Y = \sum_y y F_y$ commute.

Proof.

ii \Rightarrow i Put $M_{x,y} = E_x F_y = F_y E_x$. Since the product of commuting projections is a projector, $M = \{M_{x,y}\}$ is a (sharp) observable. Moreover, $E = f \circ M$, $F = g \circ M$, where $f(x, y) = x$, $g(x, y) = y$. Thus, E, F are compatible.

i \Rightarrow ii If E, F are compatible, there exists an observable $M = \{M_z\}$ such that

$$E_x = \sum_{z: f(z)=x} M_z, \quad F_y = \sum_{z: g(z)=y} M_z$$

for some functions f, g (we agree to set M_z to zero if there is no z satisfying the condition in question).

Lemma 2.14. *Let $0 \leq A \leq P$, where P is a projector, then $[A, P] = 0$.*

Proof. We have $(I - P)A(I - P) = 0$, hence $\sqrt{A}(I - P) = 0$, hence $A(I - P) = 0$, hence $A = AP$, hence $PA = (AP)^* = AP$. \square

Now, if $f(z) = x$, $E_x \geq M_z$ and by lemma $[E_x, M_z] = 0$. If $f(z) \neq x$, $I - E_x \geq M_z$, and again $[E_x, M_z] = 0$. Thus, E_x commute with all M_z and hence with $F_y = \sum_{z: g(z)=y} M_z$.

ii \Rightarrow iii is obvious. Conversely, let $[X, Y] = 0$. In this case, $[f(X), g(Y)] = 0$ for arbitrary polynomials f, g . Taking the polynomials, which vanish at all points of the corresponding spectra except x (resp. y), we obtain $[E_x, F_y] = 0$. \square

If E, F are compatible, then denoting $\hat{M}_{xy} = \sum_{z: f(z)=x, g(z)=y} M_z$, we have

$$E_x = \sum_y \hat{M}_{xy}, \quad F_y = \sum_x \hat{M}_{xy}.$$

Observable \hat{M} describes the statistics of joint measurement of E, F . Their joint probability distribution in the state S is given by

$$\mu_S^{EF}(x, y) = \text{Tr } S \hat{M}_{xy}.$$

One can similarly define a joint measurement and probability distribution for an arbitrary finite collection of compatible observables.

Exercise 2.15. Show that the only observable compatible with all quantum observables is a constant, i.e. the observable equal to a scalar multiple of the unit operator.

Existence of (the vast variety of) incompatible observables is a manifestation of the quantum feature of *complementarity*. Physical measurements on micro-objects

are implemented by macroscopic experimental devices, each of which requires a specific organization of the environment in space and time (such as e.g. Stern–Gerlach experiments). Different ways of organization, corresponding to measurements of different observables, may be mutually exclusive (despite the fact that they relate to the identically prepared micro-object), i.e. complementary. Complementarity is the first fundamental distinction between quantum and classical statistical models.

Example 2.16. Consider a spin-1/2 particle, and let the unit vector $\vec{a} = (a_x, a_y, a_z)$ give a direction, then

$$\sigma(\vec{a}) = a_x \sigma_x + a_y \sigma_y + a_z \sigma_z = \begin{bmatrix} a_z & a_x - i a_y \\ a_x + i a_y & -a_z \end{bmatrix} \quad (2.14)$$

is the observable of the projection of the spin² onto the direction \vec{a} . Indeed, the operator $\sigma(\vec{a})$ has the eigenvalues ± 1 (spin along and opposite the direction \vec{a}), and the spectral decomposition is given by

$$\sigma(\vec{a}) = |\psi(\vec{a})\rangle\langle\psi(\vec{a})| - |\psi(-\vec{a})\rangle\langle\psi(-\vec{a})| = \begin{bmatrix} a_z & a_x - i a_y \\ a_x + i a_y & -a_z \end{bmatrix}. \quad (2.15)$$

Recall that \vec{a} has the Euler angles (θ, ϕ) , and $\psi(\vec{a})$, given by (2.9), is the vector of the pure state in \mathcal{H} , with the spin direction \vec{a} . The corresponding density operator, which is the one-dimensional projector onto $\psi(\vec{a})$, is

$$S(\vec{a}) = |\psi(\vec{a})\rangle\langle\psi(\vec{a})| = \frac{I + \sigma(\vec{a})}{2}.$$

Exercise 2.17. Prove

$$\sigma(\vec{a}_1)\sigma(\vec{a}_2) = (\vec{a}_1 \cdot \vec{a}_2)I + i\sigma(\vec{a}_1 \times \vec{a}_2). \quad (2.16)$$

Written for the basis vectors this amounts to

$$\begin{aligned} \sigma_x^2 &= I, & \sigma_y^2 &= I, & \sigma_z^2 &= I, \\ \sigma_x \sigma_y &= i \sigma_z, & \sigma_y \sigma_z &= i \sigma_x, & \sigma_z \sigma_x &= i \sigma_y. \end{aligned} \quad (2.17)$$

Taking into account that $\text{Tr } \sigma(\vec{a}) = 0$, the relation (2.16) implies the formula for the mean value of $\sigma(\vec{b})$:

$$\text{Tr } S(\vec{a})\sigma(\vec{b}) = \vec{a} \cdot \vec{b}.$$

Another corollary of (2.16) is

$$[\sigma(\vec{a}_1), \sigma(\vec{a}_2)] = 2i\sigma(\vec{a}_1 \times \vec{a}_2), \quad (2.18)$$

which implies that $\sigma(\vec{a}_1), \sigma(\vec{a}_2)$ are compatible if and only if $\vec{a}_1 = \pm \vec{a}_2$. In particular, the spin components $\sigma_x, \sigma_y, \sigma_z$ are incompatible observables.

² In physical applications there is a dimensional factor $\hbar/2$, which can be removed by an appropriate choice of units.

2.3.3 The uncertainty relation

If X, Y are two operators, we can write

$$XY = X \circ Y + \frac{1}{2}[X, Y],$$

where $X \circ Y = \frac{1}{2}(XY + YX)$ is the symmetrized or *Jordan product* of X, Y .

Let S be some state. For an arbitrary collection $X = \{X_1, \dots, X_n\}$ of sharp, real observables we introduce two real matrices, one symmetric and another skew-symmetric

$$B_S(X) = \left[\text{Tr } S X_j^0 \circ X_k^0 \right]_{j,k=1,\dots,n}; \quad C_S(X) = \left[i \text{Tr } S [X_j, X_k] \right]_{j,k=1,\dots,n}, \quad (2.19)$$

where $X_j^0 = X_j - I E_S(X_j)$. The matrix $B_S(X)$ is called the *covariance matrix*, while $C_S(X)$ is called the *commutation matrix* of X . We have

$$B_S(X) \geq \pm \frac{i}{2} C_S(X) \quad (2.20)$$

in the sense of an inequality between complex Hermitian matrices. Indeed, the Hermitian matrix

$$B_S(X) - \frac{i}{2} C_S(X) = \left[\text{Tr } S X_j^0 X_k^0 \right]_{j,k=1,\dots,n}$$

is positive definite, since for arbitrary $c_j \in \mathbb{C}$

$$\sum_{j,k=1}^n \bar{c}_j c_k \text{Tr } S X_j^0 X_k^0 = \text{Tr } S Z^* Z \geq 0,$$

where $Z = \sum_{j=1}^n c_j X_j^0$. This produces the inequality (2.20) with the $+$ sign. The inequality with the minus sign follows after taking the transposes.

For two observables $X_1 = X - I E_S(X)$ and $X_2 = Y - I E_S(Y)$, the inequality (2.20) is equivalent to the Schrödinger–Robertson uncertainty relation

$$D_S(X) D_S(Y) \geq \{E_S(X - I E_S(X)) \circ (Y - I E_S(Y))\}^2 + \frac{1}{4} |E_S[X, Y]|^2, \quad (2.21)$$

where

$$D_S(X) = \text{Tr } S (X - I E_S(X))^2 \quad (2.22)$$

is the *variance* of the sharp, real observable X in the state S . The quantity

$$E_S(X - I E_S(X)) \circ (Y - I E_S(Y)) \quad (2.23)$$

represents the *covariance* of X, Y in the state S . If the observables X, Y are compatible, this quantity coincides with the covariance of random variables, in the sense

of probability theory, with respect to their joint probability distribution. In this case, $E_S[X, Y] = 0$ and (2.21) reduces to the Cauchy–Schwarz inequality for the covariance of random variables. However, if X, Y are not compatible, X, Y are not measurable in a single experiment and the variances $D_S(X), D_S(Y)$ in the uncertainty relation refer to the two different measurements performed over different representatives of the same statistical ensemble, while the covariance is just a formal characteristic of the state.

Exercise 2.18. Prove the *noncommutative Cauchy–Schwarz inequality*

$$|\mathrm{Tr} SX^*Y|^2 \leq \mathrm{Tr} SX^*X \mathrm{Tr} SY^*Y, \quad (2.24)$$

for arbitrary state S and operators X, Y in \mathcal{H} .

2.3.4 Convex structure of observables

The set of all (finite-valued) quantum observables in a Hilbert space \mathcal{H} will be denoted by $\mathfrak{M}(\mathcal{H})$. Given a Hilbert space \mathcal{H} , the sets $\mathfrak{S}(\mathcal{H})$ of quantum states and $\mathfrak{M}(\mathcal{H})$ of quantum observables, by construction, together with the generalized Born–von Neumann statistical rule (2.13), satisfy the axioms of separated statistical model. However, this model has an additional convexity structure in $\mathfrak{M}(\mathcal{H})$, which is to some extent similar to that of $\mathfrak{S}(\mathcal{H})$.

Let $\{M^j\}$ be a finite collection of observables with the same space of outcomes \mathcal{X} . Given a probability distribution $\{p_j\}$, we can define the *mixture* $M = \{M_x; x \in \mathcal{X}\}$ of these observables in an obvious way

$$M_x = \sum_j p_j M_x^j; \quad x \in \mathcal{X}. \quad (2.25)$$

Thus, the set $\mathfrak{M}_{\mathcal{X}}$ of all observables with a given space of outcomes \mathcal{X} is a convex set. Similar to what is the case with states, mixtures of observables describe measurements with fluctuations in the classical parameters of the measuring procedure.

The following result describes the relation between observables without classical randomness and sharp observables.

Theorem 2.19. *Any sharp observable $M \in \mathfrak{M}_{\mathcal{X}}$ is an extreme point of $\mathfrak{M}_{\mathcal{X}}$. In the reverse direction, an extreme observable $M \in \mathfrak{M}_{\mathcal{X}}$, with commuting components, $[M_x, M_{x'}] \equiv 0$, is sharp.*

Proof. Let M be sharp, and assume $M = pM^1 + (1 - p)M^2$, $0 < p < 1$. In this case, similar to (2.7)

$$pM_x^1(I - M_x^1) + (1 - p)M_x^2(I - M_x^2) + p(1 - p)(M_x^1 - M_x^2)^2 = 0. \quad (2.26)$$

and therefore $M_x^1 \equiv M_x^2 \equiv M_x$, and M is extreme.

Now notice that in all cases, $M_x \leq I$ and hence $M_x^2 \leq M_x$. Taking some $x \in \mathcal{X}$ we have

$$M_x = \frac{1}{2}M_x^2 + \frac{1}{2}(2M_x - M_x^2); \quad (2.27)$$

$$M_{x'} = \frac{1}{2}M_{x'}(I + M_x) + \frac{1}{2}M_{x'}(I - M_x), \quad x' \neq x. \quad (2.28)$$

If $[M_x, M_{x'}] \equiv 0$, these relations represent M as a convex combination (with equal weights) of two other observables. If M is extreme, they should coincide with M , in particular $M_x^2 = M_x$. Hence, M is sharp. \square

It follows that in the classical case extreme observables coincide with the sharp ones, giving them a very clear characterization as observables without randomness in the measuring procedure. In the quantum statistical model, things are not so simple (and much more interesting). We call the two-valued observables *tests* (they are also called *propositions*, *questions*, *effects* in the literature, and play a central role in various axiomatic approaches although, as we have just seen, they have a rather special property as concerns the important issue of extremity). The set of extreme quantum observables is exhausted by sharp observables only in this case of two outcomes. This fact follows from the theorem, because any test necessarily has commuting components $\{M_0, M_1 = I - M_0\}$. Thus, any extreme test is completely determined by the projector $P = M_0$.

For observables with more than two values, we consider the following construction.

Definition 2.20. Let $\{|\psi_x\rangle\}$ be an arbitrary, finite collection of not necessarily normalized vectors in \mathcal{H} such that $\sum_x |\psi_x\rangle\langle\psi_x| = I$. Such a collection is called an *overcomplete system*.

If $\{|\psi_x\rangle\}$ is an overcomplete system, an arbitrary vector $|\psi\rangle \in \mathcal{H}$ can be represented as a linear combination

$$|\psi\rangle = \sum_x c_x |\psi_x\rangle \quad \text{with} \quad c_x = \langle\psi_x|\psi\rangle, \quad (2.29)$$

where the coefficients $\{c_x\}$ need not be unique, because the vectors $|\psi_x\rangle$ may be linearly dependent. There also is a corresponding matrix representation of the operators. In a d -dimensional Hilbert space, overcomplete systems with more than d vectors always exist, and can be built from any complete (possibly, linearly dependent) system as follows. Let $\{|\phi_x\rangle\}$ be a complete system of vectors in \mathcal{H} . The corresponding *Gram operator* is defined by

$$G = \sum_x |\phi_x\rangle\langle\phi_x|. \quad (2.30)$$

Completeness implies that G is nondegenerate. Now, $\{|\psi_x\rangle\}$, with $|\psi_x\rangle = G^{-1/2}|\phi_x\rangle$, is an overcomplete system because

$$\sum_x |\psi_x\rangle\langle\psi_x| = G^{-1/2} \sum_x |\phi_x\rangle\langle\phi_x| G^{-1/2} = I. \quad (2.31)$$

An overcomplete system determines a quantum observable M with the components

$$M_x = |\psi_x\rangle\langle\psi_x|, \quad x \in \mathcal{X}. \quad (2.32)$$

In particular, for any orthonormal basis $\{e_x\}$, the observable $M = \{|e_x\rangle\langle e_x|\}$ is sharp and hence, an extreme observable.

Theorem 2.21. *An observable (2.32) is extreme if and only if the operators M_x are linearly independent in the real vector space $\mathfrak{B}_h(\mathcal{H})$.*

Proof. Let M be extreme and assume that

$$\sum_x c_x |\psi_x\rangle\langle\psi_x| = 0. \quad (2.33)$$

For $\epsilon > 0$ small enough, we define

$$M_x^\pm = (1 \pm \epsilon c_x) M_x \geq 0 \quad x \in \mathcal{X}.$$

In this case, M^\pm are observables and, by construction, $M = \frac{1}{2}M^+ + \frac{1}{2}M^-$. Therefore, $M_x^+ = M_x^- = M_x$ since M is extreme. Thus, (2.33) implies $c_x = 0$, i.e. the components of M are linearly independent.

Conversely, let

$$|\psi_x\rangle\langle\psi_x| = pM_x^1 + (1-p)M_x^2$$

be a convex decomposition of M . In this case, $0 \leq pM_x^1 \leq |\psi_x\rangle\langle\psi_x|$. By Lemma 2.14,

$$M_x^1 |\psi_x\rangle\langle\psi_x| = |\psi_x\rangle\langle\psi_x| M_x^1 = \langle\psi_x|\psi_x\rangle M_x^1,$$

whence $M_x^1 = \lambda_x |\psi_x\rangle\langle\psi_x|$, with $\lambda_x = \langle\psi_x|M_x^1|\psi_x\rangle/\langle\psi_x|\psi_x\rangle^2$. Now, $\sum_x \lambda_x |\psi_x\rangle\langle\psi_x| = I$, i.e.

$$\sum_x (\lambda_x - 1) |\psi_x\rangle\langle\psi_x| = 0.$$

Due to the linear independence of M_x , $\lambda_x = 1$, and $M_x^1 = M_x$ for all x , which means that M is extreme. \square

From the result of Exercise 1.5, it follows that the dimensionality of the real space of Hermitian operators is equal to d^2 . Thus, all overcomplete systems with n linearly independent components, $d < n \leq d^2$ determine unsharp extremal observables. A concrete example of such an observable with $n = 3$, $d = 2$ will be considered in the next section.

2.4 Statistical discrimination between quantum states

2.4.1 Formulation of the problem

In this section, we consider a statistical decision problem which will allow us to pass to the optimal information transmission via quantum channels in Part II of the book.

Let a quantum system be prepared in one of the states S_x , $x = 1, \dots, n$. The observer is allowed to perform arbitrary measurements aimed at determining, in the best possible way, in which of these states, given a priori, the system was actually prepared. This kind of decision problem is typical for mathematical statistics and its applications in communication theory, where we deal with the optimal signal detection or estimation (see Section 3.3.1). On the other hand, in high-precision experiments, researchers are already able to operate with elementary quantum systems such as single ions, atoms, and photons. This leads to potentially important applications, such as quantum communication and quantum cryptography. A quite important issue is that of extracting the maximum possible information about the state of a given quantum system. In proposals currently under discussion for application to quantum computing, the information is written into states of elementary quantum cells, named qubits, and is read via quantum measurements. From a statistical point of view, a measurement gives an estimate for the quantum state, either as a whole, or for some of its components (parameters), and the problem of finding the most informative measurement arises.

The statistic of the whole measurement procedure is described by an observable, i.e. by resolution of the identity $M = \{M_x\}$ in the system space \mathcal{H} . The probability to make a decision y , under the condition that the system was in the state S_x , is equal to $p_M(y|x) = \text{Tr } S_x M_y$. In particular, the probability of making a correct decision is equal to $p_M(x|x)$. Let us additionally assume that the states S_x appear with probabilities π_x (e.g. in the case of equiprobable states $\pi_x = 1/n$). In this case, the average probability of making a correct decision is equal to

$$\mathcal{P}\{M\} = \sum_{x=1}^n \pi_x p_M(x|x),$$

and the problem is to maximize it with respect to all possible observables M .

2.4.2 Optimal observables

Denoting by $W_x = \pi_x S_x \geq 0$, the average probability of a correct decision can be written explicitly as an affine function

$$\mathcal{P}\{M\} = \sum_{x=1}^n \text{Tr } W_x M_x,$$

on the convex set of observables with n outcomes

$$\mathfrak{M}_n = \left\{ M = \{M_x\}_{x=1,\dots,n} : M_x \geq 0, \sum_{x=1}^n M_x = I \right\}.$$

Maximization of an affine function defined on a compact convex set is a typical linear programming problem.

Theorem 2.22. *The average probability of a correct decision $\mathcal{P}\{M\}$ attains its maximum at an extreme point of the set \mathfrak{M}_n . An observable M^0 is optimal if and only if there exists a Hermitian operator Λ , such that*

- i. $(\Lambda - W_x)M_x^0 = 0$;
- ii. $\Lambda \geq W_x$.

The following duality relation holds

$$\max\{\mathcal{P}\{M\} : M \in \mathfrak{M}_n\} = \min\{\text{Tr } \Lambda : \Lambda \geq W_x, x = 1, \dots, n\}. \quad (2.34)$$

Proof. Since $\mathcal{P}\{M\}$ is a continuous affine function on a compact, convex set \mathfrak{M}_n , the first statement follows from the result of Exercise 1.7.

Let us first prove the sufficiency of the conditions i, ii. Let Λ be a Hermitian operator satisfying the conditions i, ii. By using these conditions and the property (1.13), we obtain

$$\mathcal{P}\{M\} = \text{Tr} \sum_x W_x M_x \leq \text{Tr} \sum_x \Lambda M_x = \text{Tr } \Lambda. \quad (2.35)$$

From i, we have $\Lambda M_x^0 = W_x M_x^0$. If we sum over x and take the trace of the resulting equation, we obtain

$$\text{Tr } \Lambda = \text{Tr } \Lambda \sum_x M_x^0 = \text{Tr} \sum_x W_x M_x^0 = \mathcal{P}\{M^0\}. \quad (2.36)$$

Hence, we obtain

$$\mathcal{P}\{M\} \leq \mathcal{P}\{M^0\}, \quad \text{for all } M, \quad (2.37)$$

and therefore M^0 maximizes \mathcal{P} . Notice that

$$\Lambda = \sum_x W_x M_x^0 = \sum_x M_x^0 W_x.$$

Now, we prove the necessity. Put $M_x = A_x^2$, where A_x are Hermitian operators satisfying the condition $\sum_x A_x^2 = I$. Applying the method of Lagrange multipliers, we can reduce the problem of maximizing $\mathcal{P}\{M\}$ on the set \mathfrak{M}_n to the problem of maximizing the function

$$\text{Tr} \sum_x W_x A_x^2 - \text{Tr } \Lambda \left(\sum_x A_x^2 - I \right), \quad (2.38)$$

where Λ is a Hermitian operator, over all collections of Hermitian operators A_x . Here, the matrix elements of the Hermitian operator Λ represent the array of the Lagrange multipliers for this problem. Let A_x^0 be the optimal collection. In addition, let $A_x = A_x^0 + \epsilon Y_x$, with Hermitian Y_x , and consider (2.38) as a function of ϵ . By considering the coefficients of ϵ and ϵ^2 , we obtain the conditions

$$\begin{aligned}\text{Tr}[(W_x - \Lambda)A_x^0 + A_x^0(W_x - \Lambda)]Y_x &= 0, \\ \text{Tr}(W_x - \Lambda)Y_x^2 &\leq 0\end{aligned}$$

for an arbitrary Hermitian Y_x , i.e.

$$(W_x - \Lambda)A_x^0 + A_x^0(W_x - \Lambda) = 0, \quad \Lambda - W_x \geq 0.$$

The second inequality is just the condition ii. of the theorem. By putting $M_x^0 = (A_x^0)^2$, the first relation implies $\text{Tr}(\Lambda - W_x)M_x^0 = 0$, which, together with the second inequality, leads to the condition i. \square

Exercise 2.23. Prove that the operator Lagrange multiplier Λ is the unique solution Λ^0 of the dual problem on the right side of (2.34).

Let us illustrate the meaning and applications of these conditions with some examples.

Example 2.24. Consider the classical case where all density operators S_x , and hence the operators W_x , commute. In this case, there is a common orthonormal basis $\{|\omega\rangle\}$ in which they are all diagonal,

$$W_x = \sum_{\omega} W_x(\omega) |\omega\rangle \langle \omega|.$$

In this case, the dual problem has the unique solution

$$\Lambda^0 = \sum_{\omega} \max_x W_x(\omega) |\omega\rangle \langle \omega|,$$

where $\max_x W_x(\omega)$ is the upper envelope of the family of functions $W_x(\omega)$; $x = 1, \dots, n$. A solution of the initial problem is given by the formula

$$M_x^0 = \sum_{\omega} \mathbf{1}_{\Omega_x}(\omega) |\omega\rangle \langle \omega|,$$

where $\mathbf{1}_{\Omega_x}$ denotes the indicator of the subset Ω_x , where subsets $\Omega_x \subseteq \{\omega : \Lambda^0(\omega) = W_x(\omega)\}$ are assumed to be non-intersecting and to constitute a partition of the set $\Omega = \{\omega\}$.

This solution amounts to the principle of *maximal likelihood* in classical statistics. The decision x should be taken for those observations ω for which the posterior gain

$W_x(\omega)$ is maximal. Thus, in the classical case, the optimal observable can always be chosen to be sharp. This is directly related to the fact that in the commutative case extreme points of the set \mathfrak{M}_n are precisely the orthogonal resolutions of the identity (cf. Theorem 2.19).

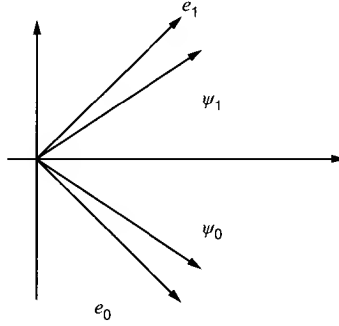


Figure 2.2. Discrimination between two pure states.

Example 2.25 (Discrimination between the two states). Let S_0, S_1 be two density operators, with π_0, π_1 their a priori probabilities. An observable (test) is given by resolution of the identity $\{M_0, M_1\}$, so that $M_0 + M_1 = I$. This case can be reduced to the previous one, by representing

$$W_0 = \pi_1 S_1 + W'_0, \quad W_1 = \pi_1 S_1 + W'_1,$$

where the operators $W'_0 = \pi_0 S_0 - \pi_1 S_1$, $W'_1 = 0$ commute. Thus

$$\mathcal{P}\{M\} = \pi_1 + \text{Tr} (W'_0 M_0 + W'_1 M_1), \quad (2.39)$$

and

$$\Lambda^0 = \max\{\pi_0 S_0 - \pi_1 S_1, 0\} = (\pi_0 S_0 - \pi_1 S_1)_+. \quad (2.40)$$

Any optimal M_0^0 has the form

$$M_0^0 = \mathbf{1}_{(0,\infty)}(\pi_0 S_0 - \pi_1 S_1) + X_0, \quad (2.41)$$

where the first term is the projector onto the positive part of the spectrum of $\pi_0 S_0 - \pi_1 S_1$, and the second is an operator supported by the null subspace of $\pi_0 S_0 - \pi_1 S_1$, such that $0 \leq X_0 \leq I$. The maximum is given by

$$\mathcal{P}\{M\} = \pi_1 + \text{Tr} (\pi_0 S_0 - \pi_1 S_1)_+. \quad (2.42)$$

By using the relation

$$(\pi_0 S_0 - \pi_1 S_1)_+ = \frac{1}{2} |\pi_0 S_0 - \pi_1 S_1| + \frac{1}{2} (\pi_0 S_0 - \pi_1 S_1),$$

we obtain

$$\max \mathcal{P}\{M\} = \frac{1}{2} (1 + \|\pi_0 S_0 - \pi_1 S_1\|_1).$$

In particular, taking $\pi_0 = \pi_1 = \frac{1}{2}$, we have

$$\max \mathcal{P}\{M\} = \frac{1}{2} \left(1 + \frac{1}{2} \|S_0 - S_1\|_1 \right).$$

so that the more distinguishable the density operators S_0 and S_1 are in the sense of the trace norm, the greater the maximum will be.

Proposition 2.26. *Let $S_0 = |\psi_0\rangle\langle\psi_0|$, $S_1 = |\psi_1\rangle\langle\psi_1|$ be pure states. In this case, the maximum of $\mathcal{P}\{M\}$ is given by*

$$\max_M \mathcal{P}\{M\} = \frac{1}{2} \left(1 + \sqrt{1 - 4\pi_0\pi_1|\langle\psi_0|\psi_1\rangle|^2} \right), \quad (2.43)$$

and is achieved on a sharp observable. In particular, if $\pi_0 = \pi_1 = 1/2$, the maximum is equal to

$$\max_M \mathcal{P}\{M\} = \frac{1}{2} \left(1 + \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^2} \right). \quad (2.44)$$

Proof. Consider the eigenvalue problem for the rank-two operator

$$\pi_0 |\psi_0\rangle\langle\psi_0| - \pi_1 |\psi_1\rangle\langle\psi_1|.$$

This operator may have eigenvalue zero, corresponding to the orthogonal complement to its support, which is just the linear span \mathcal{L} of the vectors $|\psi_0\rangle, |\psi_1\rangle$. The relevant part of the optimal observable is also supported by \mathcal{L} , so we can restrict ourselves to the eigenvectors that are linear combinations

$$|\psi\rangle = c_0 |\psi_0\rangle + c_1 |\psi_1\rangle.$$

Substituting this into the eigenvector equation results in

$$\pi_0 (c_0 + \langle\psi_0|\psi_1\rangle c_1) = \lambda c_0; \quad (2.45)$$

$$-\pi_1 (\langle\psi_1|\psi_0\rangle c_0 + c_1) = \lambda c_1, \quad (2.46)$$

whence the eigenvalues are

$$\lambda_{0,1} = \frac{1}{2} \left[\pi_0 - \pi_1 \pm \sqrt{1 - 4\pi_0\pi_1|\langle\psi_0|\psi_1\rangle|^2} \right]$$

with the corresponding orthonormal basis of eigenvectors $|e_0\rangle, |e_1\rangle$. Now,

$$\|\pi_0|\psi_0\rangle\langle\psi_0| - \pi_1|\psi_1\rangle\langle\psi_1|\|_1 = \lambda_0 - \lambda_1 = \sqrt{1 - 4\pi_0\pi_1|\langle\psi_0|\psi_1\rangle|^2},$$

and we get (2.43). \square

The sharp optimal observable in the subspace \mathcal{L} is just $\{|e_0\rangle\langle e_0|, |e_1\rangle\langle e_1|\}$. If $\pi_0 = \pi_1 = 1/2$, the optimal basis is situated symmetrically with respect to $|\psi_0\rangle, |\psi_1\rangle$ (see Figure 2.2).

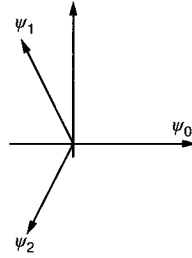


Figure 2.3. Trine on the plane.

Example 2.27. On a plane (considered as a real subspace of two-dimensional complex space), consider an “equiangular” configuration of the three vectors (see Figure 2.3)

$$|\psi_j\rangle = \begin{bmatrix} \cos \frac{2j\pi}{3} \\ \sin \frac{2j\pi}{3} \end{bmatrix}, \quad j = 0, 1, 2. \quad (2.47)$$

The corresponding density matrices $S_j = |\psi_j\rangle\langle\psi_j|$ describe the states of a two-level quantum system such as a linearly polarized photon or a spin 1/2 particle.

We have

$$S_j = \frac{1}{2} \left(I + \begin{bmatrix} \cos \frac{4j\pi}{3} & \sin \frac{4j\pi}{3} \\ \sin \frac{4j\pi}{3} & -\cos \frac{4j\pi}{3} \end{bmatrix} \right), \quad (2.48)$$

so that

$$\sum_{j=0}^2 S_j = \frac{3}{2} I$$

because $\sum_{j=0}^2 e^{i\frac{4\pi j}{3}} = 0$.

It follows that $M_k^0 = \frac{2}{3}S_k$; $k = 0, 1, 2$, is the resolution of the identity that corresponds to the overcomplete system $\{\sqrt{\frac{2}{3}}|\psi_k\rangle; k = 0, 1, 2\}$. The corresponding observable is extremal, since it satisfies the conditions of Theorem 2.21.

Let us show that, in the case of equiprobable states $\pi_j = 1/3$, observable $\{M_k^0\}$ is optimal, by checking the conditions of Theorem 2.22. Since $S_j^2 = S_j$,

$$\Lambda^0 = \sum_{j=0}^2 \frac{1}{3} S_j M_j^0 = \frac{2}{9} \sum_{j=0}^2 S_j = \frac{1}{3} I.$$

so that condition ii. is satisfied: $I/3 = \Lambda^0 \geq S_j/3$. Condition i. is also satisfied since

$$\left(\Lambda^0 - \frac{1}{3}S_j\right)M_j^0 = \frac{1}{3}(I - S_j)S_j = 0.$$

Thus, $\max \mathcal{P}\{M\} = \text{Tr } \Lambda^0 = 2/3$. Let us now find the maximum over all sharp observables with three outcomes. A nontrivial orthogonal resolution of the identity with three components in the two-dimensional space has the form $M_0 = |e_0\rangle\langle e_0|$, $M_1 = |e_1\rangle\langle e_1|$, $M_2 = 0$, where $|e_0\rangle, |e_1\rangle$, is an arbitrary basis. Thus, the problem reduces to the optimal discrimination of the two equiprobable states S_0, S_1 . Applying relation (2.44) to the case $\langle\psi_0|\psi_1\rangle = -\frac{1}{2}$, we obtain

$$\max_{M\text{-sharp}} \mathcal{P}\{M\} = \frac{2}{3} \frac{1 + \sqrt{3}/2}{2} < \frac{2}{3} = \max_{M \in \mathfrak{M}} \mathcal{P}\{M\}.$$

Thus, in the quantum state discrimination, using unsharp observables can lead to an improvement in comparison with the sharp ones. Let us stress that in the similar classical problem (of discrimination between probability distributions) unsharpness or randomization can never lead to such an improvement. From a geometrical point of view, the reason for this phenomenon is, of course, that in the quantum case not all extremal observables (among which the optimal one is found) are sharp ones.

2.5 Notes and references

1. Among the texts devoted to the mathematical foundations of Quantum Mechanics let us mention the monographs of von Neumann [212], Mackey [153], Segal [181], Faddeev, and Yakubovsky [57]. A survey of axiomatic approaches is given in the paper of Wightman [219] and also in the Comments to Ch. VIII of the book of Reed and Simon [171]. One aim of any axiomatics that is longed for is a derivation of the Hilbert space formalism. Such a derivation within the framework of the operational approach (see below) was realized, in the finite dimensional case, in the paper of

Araki [7] (for a more recent approach, see Hardy [75]). However, this is a separate topic; for our purposes it will suffice to postulate the state space $\mathfrak{S} = \mathfrak{S}(\mathcal{H})$.

The description of mixed states by density operators (matrices) was introduced independently by von Neumann, Weyl, and Landau (see footnote 172 in Section IV.2 of von Neumann's book [212]).

The axiomatic approach, in which the central role is played by the duality between the two partially ordered linear spaces, generated by a convex set of quantum states, and by the order interval $[0, I]$ of two-valued observables (effects or tests), was developed by the school of Ludwig [152], [61], see also Davies [48] (this approach is usually called *operational*). A detailed investigation of the statistical structure of quantum theory, based on the operational approach, is undertaken in the books [107], [99], where one can find a detailed bibliography.

In Proposition 2.1, we had to assume the existence of the maximal observable for the sole reason that, for simplicity, from the beginning we restricted ourselves to observables with finite sets of outcomes. With an appropriate generalization the existence of the maximal observable (in general, with an infinite Ω) can be deduced from compatibility of all observables of the model. Essentially, this is the message of the famous Kolmogorov's extension theorem for a random process defined by system of compatible finite-dimensional distributions. A detailed study of the notion of *stochastic compatibility*, including the proof that the separated model, in which all observables are stochastically compatible, can be embedded within the Wald model, can be found in the paper [96].

A brief introduction to the basic notions of quantum theory targeted at quantum information science is given in the books of Nielsen and Chuang [158] and Hayashi [78]. The papers of Fuchs [63], D'Ariano [40], Masanes, and Müller [156] elaborate on information-theoretical approaches to the foundations of quantum theory.

2. The generalized concept of a quantum observable presented here was developed by Ludwig [152], Davies and Lewis [48], and Holevo [107]. As concerns the convex structure of the set of observables, see Kraus [139]. In analysis and in signal theory, overcomplete systems are known under the name “(rigid) frames”.

The spin of a quantum particle is intrinsically related to representations of the group of rotations of three dimensional Euclidean space [218], [59], [153]. Results related to Theorem 2.21 are discussed in the paper of Davies [47].

3. The problem of discriminating the two pure quantum states was treated, in 1968, in the paper of Bakut and Shchurov [14]. Such problems arise naturally in connection with the detection of weak light sources, and in quantum optics [86]. A physical device that achieves the bound (2.43) for two coherent states of a radiation field (see Chapter 12), called *Dolinar receiver*, was described in [86]. It uses photon counting as a measurement, together with an ingenious feedback from the output of the photon detector to the input field of the counter.

General detection and estimation theory for quantum states was developed in the 1970s in works of Helstrom [85], Holevo [91], Yuen and Lax [230], Stratono-
vich [204], Belavkin [17], and other authors. Example 2.27, with the three equiangu-
lar states, was proposed in [94]. For further details and extensive references, see the
monographs [86], [107]. The new interest in quantum estimation theory emerged in
connection with the ideas of quantum computation. Any such computation is com-
pleted by measuring parameters of the final state of the quantum computer, which
should be maximally precise. The achievements of modern quantum estimation the-
ory, including asymptotic theory, are presented in the books of Hayashi [78] and
Petz [168]. The Bayes problem is a linear programming problem (see. [154], [172]),
and Theorem 2.22 could be proved, basing on the corresponding duality theorem. An
interesting variation of this problem proposed by Peres [165] concerns the *unambigu-
ous discrimination* between quantum states when an inconclusive result is allowed.
For a survey of the recent status of the field see e.g. Herzog [87].

Chapter 3

Composite systems and entanglement

3.1 Composite systems

3.1.1 Tensor products

The new possibilities that are offered by quantum information are to a large extent due to the unusual features of composite quantum systems. Let \mathcal{H}_j , $j = 1, 2$ be the Hilbert spaces of two finite quantum systems with inner products $\langle \cdot | \cdot \rangle_j$. The combination of the two systems is described by the tensor product of the Hilbert spaces, which is defined as follows.

According to Section 1.1, the element $\psi \in \mathcal{H}$ defines the antilinear function $\psi(\phi) = \langle \phi | \psi \rangle$ of the argument $\phi \in \mathcal{H}$. For any two elements $\psi_j \in \mathcal{H}_j$; $j = 1, 2$, we denote by $\psi_1 \otimes \psi_2$ the bi-antilinear function of the arguments $\phi_1 \in \mathcal{H}_1, \phi_2 \in \mathcal{H}_2$, defined by the relation

$$(\psi_1 \otimes \psi_2)(\phi_1, \phi_2) = \langle \phi_1 | \psi_1 \rangle_1 \langle \phi_2 | \psi_2 \rangle_2.$$

Consider the vector space \mathcal{L} of finite linear combinations of such functions $\sum_j c_j \psi_1^j \otimes \psi_2^j$. Introduce the inner product in \mathcal{L} by letting

$$\langle \varphi_1 \otimes \varphi_2 | \psi_1 \otimes \psi_2 \rangle = \langle \varphi_1 | \psi_1 \rangle_1 \langle \varphi_2 | \psi_2 \rangle_2$$

on the generating elements $|\psi_1 \otimes \psi_2\rangle \equiv |\psi_1\rangle \otimes |\psi_2\rangle$ and extending, by linearity, to \mathcal{L} .

Exercise 3.1. Show that such a linear extension is possible, and that it is unique and satisfies all the properties of an inner product.

Definition 3.2. The space \mathcal{L} with the inner product defined above is called the *tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$.

Exercise 3.3. Prove the following statement: if $\{e_1^j\}, \{e_2^k\}$ are orthonormal bases in $\mathcal{H}_1, \mathcal{H}_2$, then $\{e_1^j \otimes e_2^k\}$ is an orthonormal basis in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\dim \mathcal{H}_1 \otimes \mathcal{H}_2 = \dim \mathcal{H}_1 \times \dim \mathcal{H}_2$.

It follows that any vector $\psi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is uniquely represented in the form

$$|\psi\rangle = \sum_{j,k} c_{jk} |e_1^j\rangle \otimes |e_2^k\rangle,$$

where c_{jk} are complex numbers. Summing with respect to j and denoting $|\psi_k\rangle = \sum_{j=1}^{d_1} c_{jk} |e_1^j\rangle \in \mathcal{H}_1$, we obtain

$$|\psi\rangle = \sum_{k=1}^{d_2} |\psi_k\rangle \otimes |e_2^k\rangle,$$

which means that the space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is isomorphic to the *direct orthogonal sum* $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_1$ of $d_2 = \dim \mathcal{H}_2$ copies of the space \mathcal{H}_1 .

For the operators X_j in the spaces \mathcal{H}_j we define the tensor product by putting

$$(X_1 \otimes X_2)(\psi_1 \otimes \psi_2) = X_1 \psi_1 \otimes X_2 \psi_2,$$

and extending by linearity. Let a basis in \mathcal{H}_2 be fixed, so that $\mathcal{H}_1 \otimes \mathcal{H}_2$ is realized as $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_1$. In this case, $X_1 \otimes X_2$ is realized as the block matrix $[X_1 x_2^{jk}]$, where $[x_2^{jk}]$ is the matrix of the operator X_2 . Now, a general operator X in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is given by a block matrix $[X_{jk}]$ whose entries are operators in \mathcal{H}_1 .

Let us recall that in Section 1.4 we denoted by $\mathcal{B}_h(\mathcal{H})$ the real linear space of all Hermitian operators (i.e. sharp real observables) in \mathcal{H} . From the result of Exercise 1.5, in the case of complex Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ we have

$$\dim \mathcal{B}_h(\mathcal{H}_1 \otimes \mathcal{H}_2) = \dim \mathcal{B}_h(\mathcal{H}_1) \cdot \dim \mathcal{B}_h(\mathcal{H}_2)$$

while in the case of real Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ the dimensionalities of spaces of symmetric operators are related by the inequality $>^1$.

Exercise 3.4. Prove the following statement: if S_j are the density operators in \mathcal{H}_j , then $S_1 \otimes S_2$ is a density operator in $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Definition 3.5. Let T be an operator in $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. The *partial trace* of T (with respect to \mathcal{H}_2) is defined as an operator $\text{Tr}_{\mathcal{H}_2} T$ in \mathcal{H}_1 , satisfying

$$\langle \phi | (\text{Tr}_{\mathcal{H}_2} T) | \psi \rangle = \sum_k \langle \phi \otimes e_2^k | T | \psi \otimes e_2^k \rangle, \quad \phi, \psi \in \mathcal{H}_1.$$

If S_{12} is a density operator corresponding to a state of the composite system, $S_1 = \text{Tr}_{\mathcal{H}_2} S_{12}$ (and the similarly defined S_2) are called *partial states* of the first or second subsystem, resp.

Partial states are the noncommutative analog of marginal probability distributions in probability theory.

¹ It is possible to consider quaternionic Hilbert spaces, in which case the inequality changes to $<$, see [7].

Exercise 3.6. Show that this definition does not depend on the choice of the orthonormal basis $\{e_2^k\}$. If $T = T_1 \otimes T_2$ then $\text{Tr}_{\mathcal{H}_2}(T_1 \otimes T_2) = (\text{Tr } T_2)T_1$.

Let $\mathcal{H}_1 \otimes \mathcal{H}_2$ be realized as $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_1$, so that $T = [T_{jk}]$, then $\text{Tr}_{\mathcal{H}_2} T = \sum_{j=1}^{d_2} T_{jj}$ and $\text{Tr}_{\mathcal{H}_1} T = [\text{Tr } T_{jk}]$.

3.1.2 Naimark's dilation

The fundamental relation between orthogonal and nonorthogonal resolutions of the identity is established by the following theorem.

Theorem 3.7 (Naimark). *Let $\{M_x\}$ be a resolution of the identity in \mathcal{H} with m outcomes, $\dim \mathcal{H} = d$. In this case, there exist a Hilbert space \mathcal{K} of dimensionality $\dim \mathcal{K} \leq md$, an isometry $V : \mathcal{H} \rightarrow \mathcal{K}$, and an orthogonal resolution of the identity $\{E_x\}$ in \mathcal{K} such that*

$$M_x = V^* E_x V, \quad x \in \mathcal{X}. \quad (3.1)$$

An *isometry* is a linear map that preserves the norms and hence the inner products of vectors in the Hilbert spaces. For arbitrary $|\phi\rangle, |\psi\rangle \in \mathcal{H}$, $\langle \phi | V^* V | \psi \rangle = \langle \phi | \psi \rangle$ holds, i.e. $V^* V = I$. The isometric embedding V allows us to identify \mathcal{H} with the subspace $V\mathcal{H}$ of the space \mathcal{K} and consider $\mathcal{H} \subset \mathcal{K}$. Then M_x can be considered as a restriction of E_x onto \mathcal{H} :

$$E_x = \begin{bmatrix} M_x & \cdots \\ \cdots & \cdots \end{bmatrix}.$$

Proof. The construction of \mathcal{K} is done in two steps. First, we define vectors $|\Psi\rangle \in \mathcal{H}_m$ with

$$|\Psi\rangle = \begin{bmatrix} |\psi_1\rangle \\ \vdots \\ |\psi_m\rangle \end{bmatrix}, \quad |\psi_x\rangle \in \mathcal{H}. \quad (3.2)$$

In addition, we define the *pre-inner product*, i.e. a form satisfying all the properties of the inner product except nondegeneracy, by the relation

$$\langle \Psi' | \Psi \rangle = \sum_x \langle \psi'_x | M_x | \psi_x \rangle. \quad (3.3)$$

The properties follow from the definition of the resolution of the identity.

In order to ensure that the norm induced by this pre-inner product is non-degenerate, one considers the quotient space $\mathcal{K} = \mathcal{H}_m / \mathcal{H}_0$, where

$$\mathcal{H}_0 = \{\Psi_0 \in \mathcal{H}_m : \langle \Psi_0 | \Psi_0 \rangle = 0\},$$

i.e. we identify the vectors whose difference has zero norm. Now, the map $V : \mathcal{H} \rightarrow \mathcal{H}_m$, defined as

$$V|\psi\rangle = \begin{bmatrix} |\psi\rangle \\ \vdots \\ |\psi\rangle \end{bmatrix}, \quad (3.4)$$

is an isometry because

$$\langle\psi|V^*V|\psi\rangle = \sum_{x=1}^m \langle\psi|M_x|\psi\rangle = \langle\psi|\psi\rangle. \quad (3.5)$$

For the orthogonal resolution of the identity $\{E_x\}$ in \mathcal{K} , we introduce $E_x|\Psi\rangle = [0, \dots, |\psi_x\rangle, \dots, 0]^T$, where the only non-zero component is on the x -th place. Hence,

$$\langle V\phi|E_x|V\psi\rangle = \langle\phi|M_x|\psi\rangle, \quad \phi, \psi \in \mathcal{H},$$

which concludes the proof. \square

We now consider an important corollary of Naimark's Dilation Theorem, which provides us with a statistical interpretation of an arbitrary resolution of the identity and establishes that the generalized definition of a quantum observable is just an extension of the standard one.

Corollary 3.8. *Let $\{M_x\}$ be an observable in \mathcal{H} . In this case, there exist a Hilbert space \mathcal{H}_0 , a unit vector $\psi_0 \in \mathcal{H}_0$ and a sharp observable $\{E_x\}$ in $\mathcal{H} \otimes \mathcal{H}_0$, such that*

$$M_x = \text{Tr}_{\mathcal{H}_0}(I \otimes |\psi_0\rangle\langle\psi_0|)E_x. \quad (3.6)$$

Proof. According to Naimark's Theorem, $M_x = V^*E_xV$, where $V : \mathcal{H} \rightarrow \mathcal{K}$ is an isometric embedding. Let us identify \mathcal{H} with the subspace $V\mathcal{H} \subseteq \mathcal{K}$. By extending, if necessary, the space \mathcal{K} , we can take $\dim \mathcal{K} = \dim \mathcal{H} \cdot d_0$, and hence

$$\mathcal{K} \simeq \mathcal{H} \oplus \dots \oplus \mathcal{H} \simeq \mathcal{H} \otimes \mathcal{H}_0,$$

where $\mathcal{H}_0 = \mathbb{C}^{d_0}$ is the coordinate complex Hilbert space of dimensionality d_0 , and \mathcal{H} is identified with the first term in the direct sum, or with the subspace $\mathcal{H} \otimes |\psi_0\rangle$, where $|\psi_0\rangle = [1, 0, \dots, 0]^T$. In this case,

$$(I \otimes |\psi_0\rangle\langle\psi_0|)E_x = \begin{bmatrix} M_x & \dots \\ 0 & 0 \end{bmatrix},$$

so that (3.6) indeed holds. \square

Thus, every observable can be realized as a sharp observable in the composite system by introducing an auxiliary system in a fixed state $S_0 = |\psi_0\rangle\langle\psi_0|$. Such a way of realization can be called *quantum randomization*.

In classical statistics, randomization, i.e. the introduction of a “roulette”, can be a useful trick, but it can never increase the information about the state of the observed system. From the result of Section 2.4.2 it follows that in quantum statistics this is no longer true. Paradoxically, quantum randomization allows us to extract more information about the state of the observed system than is contained in the sharp observables when we are not using an auxiliary system.

3.1.3 Schmidt decomposition and purification

Consider a state S_{12} in the Hilbert space of the composite system $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Definition 3.9. A pure state S_{12} is called *entangled* if it cannot be represented as a product state $S_1 \otimes S_2$.

Thus, any unit vector $\psi_{12} \in \mathcal{H}_1 \otimes \mathcal{H}_2$ that is a nontrivial superposition of product vectors generates a pure entangled state. An important example is the *maximally entangled state*, generated by the vector

$$|\psi_{12}\rangle = \frac{1}{\sqrt{d}} \sum_{j=1}^d |e_j^1\rangle \otimes |e_j^2\rangle \quad (3.7)$$

in the space $\mathcal{H}_1 \otimes \mathcal{H}_2$, where $d = \dim \mathcal{H}_1 = \dim \mathcal{H}_2$ and $\{e_j^{1,2}\}$ are orthonormal bases in $\mathcal{H}_{1,2}$. The meaning of this name will be clarified later in Section 7.5. Meanwhile, notice that the partial states of the maximally entangled state are the chaotic states in \mathcal{H}_1 and \mathcal{H}_2 :

$$\text{Tr}_{\mathcal{H}_2} |\psi_{12}\rangle\langle\psi_{12}| = \frac{I_1}{d}$$

and similarly for \mathcal{H}_2 .

We shall often make use of the following result:

Theorem 3.10 (Schmidt Decomposition). *Let $S_{12} = |\psi\rangle\langle\psi|$ be a pure state in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $S_1 = \text{Tr}_{\mathcal{H}_2} S_{12}$, $S_2 = \text{Tr}_{\mathcal{H}_1} S_{12}$ be its partial states. In this case, the density operators S_1 , S_2 have the same non-zero eigenvalues λ_j . Furthermore,*

$$|\psi\rangle = \sum_j \sqrt{\lambda_j} |e_j^1\rangle \otimes |e_j^2\rangle \quad (3.8)$$

where $\{e_j^{1,2}\}$ are orthonormal eigenvectors of S_1 , resp. of S_2 .

Proof. Taking the orthonormal basis $\{e_j^1\}$ of eigenvectors in \mathcal{H}_1 , we can write $|\psi\rangle$ as

$$|\psi\rangle = \sum_j |e_j^1\rangle \otimes |h_j^2\rangle, \quad (3.9)$$

with some vectors $|h_j^2\rangle \in \mathcal{H}_2$. Computation of the partial trace provides us with

$$\sum_{j,k} \langle h_j^2 | h_k^2 \rangle |e_k^1\rangle \langle e_j^1| = \sum_j \lambda_j |e_j^1\rangle \langle e_j^1| \equiv S_1, \quad (3.10)$$

and therefore $\langle h_j^2 | h_k^2 \rangle = \lambda_j \delta_{jk}$ (Kronecker's delta). Thus, we can put $|e_j^2\rangle = \frac{1}{\sqrt{\lambda_j}} |h_j^2\rangle$ for $\lambda_j > 0$, and complete this orthonormal system to a basis in \mathcal{H}_2 , consisting of eigenvectors of the operator S_2 . \square

Conversely, an arbitrary mixed state of a quantum system can be “purified” i.e. extended to a pure state of a larger system:

Theorem 3.11 (Purification). *Let S_1 be a state in \mathcal{H}_1 . In this case, there exist a reference Hilbert space \mathcal{H}_2 of the same dimensionality as \mathcal{H}_1 and a pure state $|\psi\rangle\langle\psi|$ in $\mathcal{H}_1 \otimes \mathcal{H}_2$ such that $S_1 = \text{Tr}_{\mathcal{H}_2} |\psi\rangle\langle\psi|$.*

For any other pure state $|\psi'\rangle\langle\psi'|$ in $\mathcal{H}_1 \otimes \mathcal{H}_2$ that has this property, there is a unitary operator U_2 in \mathcal{H}_2 such that $|\psi'\rangle = (I_1 \otimes U_2)|\psi\rangle$.

Proof. Diagonalize S_1 and take $|\psi\rangle$ as in (3.8), with an arbitrary basis $\{e_j^2\}$ in a Hilbert space \mathcal{H}_2 isomorphic to \mathcal{H}_1 . Any other $|\psi'\rangle$ has a decomposition (3.8) with a different basis in \mathcal{H}_2 , and any two bases in the same space are related by a unitary transformation. \square

Corollary 3.12. *Let S_{12} be a state in $\mathcal{H}_1 \otimes \mathcal{H}_2$ such that the partial state S_1 is pure. In this case, $S_{12} = S_1 \otimes S_2$.*

Proof. For a pure state S_{12} the statement obviously follows from Theorem 3.10. For an arbitrary state S_{12} , we can apply the same argument to its purification. \square

From Theorem 3.10, it also follows that the purification is essentially unique. Any two purifications with reference spaces $\mathcal{H}_2, \mathcal{H}_2'$, such that $\dim \mathcal{H}_2 \leq \dim \mathcal{H}_2'$, are related by an isometric embedding of \mathcal{H}_2 into \mathcal{H}_2' .

Remarkably, there is one universal Hilbert space that contains purifications of all states in \mathcal{H} . Consider the Hilbert space $L^2(\mathcal{H})$ of all operators X, Y, \dots in \mathcal{H} , equipped with the inner product

$$(X, Y) = \text{Tr } X^* Y.$$

Then the linear correspondence

$$|\psi\rangle\langle\varphi| \leftrightarrow |\psi\rangle \otimes \langle\varphi|, \quad (3.11)$$

where $\langle\varphi| \in \mathcal{H}^*$ (the dual space of linear functions on \mathcal{H}) is uniquely extended to the isometry between the Hilbert spaces $L^2(\mathcal{H})$ and $\mathcal{H} \otimes \mathcal{H}^*$. Denoting by L_A (resp. R_B) the operation of left multiplication by A (resp. of right multiplication by B), we have for $X = |\psi\rangle\langle\varphi|$

$$L_A R_B X \equiv AXB \leftrightarrow |A\psi\rangle \otimes \langle B^*\varphi|.$$

For an arbitrary density operator S in \mathcal{H} consider the unit vector $\sqrt{S}W \in L^2(\mathcal{H})$, where W is arbitrary unitary in \mathcal{H} . Then

$$\left(\sqrt{S}W, L_A R_B \sqrt{S}W\right) = \text{Tr } \sqrt{S}A\sqrt{S}WBW^*,$$

whence, letting $B = I$,

$$\text{Tr } SA = \left(\sqrt{S}W, (A \otimes I_{\mathcal{H}^*})\sqrt{S}W\right).$$

Therefore $\sqrt{S}W$ is a purification of S in $L^2(\mathcal{H})$ identified with $\mathcal{H} \otimes \mathcal{H}^*$ via the correspondence (3.11). Moreover, all purifications of S are obtained in this way.

Notice that purification is a mathematical construct that does not correspond to any physical operation. Indeed, it would imply cloning the quantum state S_1 .

Definition 3.13. A general mixed state $S_{12} \in \mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is called *separable (un-entangled)* if it belongs to the convex hull of the set of all product states, i.e. if it has the form of the convex combination

$$S_{12} = \sum_x \pi_x S_1^x \otimes S_2^x,$$

where $S_j^x \in \mathfrak{S}(\mathcal{H}_j)$; $j = 1, 2$ and $\{\pi_x\}$ is a finite probability distribution. The states that are not separable are called *entangled*.

A simple necessary condition for separability is that the state S_{12} has a *positive partial transpose* (PPT). An arbitrary operator in $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be represented as $A_{12} = \sum_j A_1^j \otimes A_2^j$. Its partial transpose with respect to the second system is correctly defined as

$$A_{12}^{\top_2} = \sum_j A_1^j \otimes [A_2^j]^{\top_2},$$

where $^{\top_2}$ is a transposition in the second system. In general, this condition is not sufficient, so there exist PPT entangled states.

3.2 Quantum entanglement vs “local realism”

3.2.1 Paradox of Einstein–Podolski–Rosen and Bell’s inequalities

Entanglement, along with complementarity, constitute the principal structural difference between the classical and the quantum description of systems.

The pure state of a classical system is described by a point in the phase space (practically, this is a complete set of parameter values describing “inner properties” of the system). If the classical system under consideration consists of several subsystems, an arbitrary pure state is necessarily a product of the pure states of the subsystems. By fixing the values of the parameters of the system we *ipso facto* fix the values of the parameters of all its subsystems. In this case, there is no need to involve a statistical treatment.

For quantum systems the situation is different. As Corollary 3.12 shows, for any pure entangled state S_{12} of a composite system, the partial states $S_{1,2}$ are necessarily mixed, i.e. they require a statistical description.

Such behavior is quite unusual from a classical point of view and, as will be shown in this section, is entirely incompatible with the classical mode of description (which is often called “realism”).

The key example of the unusual behavior of composite quantum systems was provided by Einstein, Podolski and Rosen (EPR), in 1935. In the 1950s Bohm presented it in a clearer form, using spin degrees of freedom instead of spatial ones, and in the 1960s it was substantially clarified by J. S. Bell, who suggested a fundamental inequality, which holds for classical composite systems satisfying a natural requirement of “locality” (separability), but is violated in quantum systems and which, in principle, can be verified experimentally.

Consider two particles with spin $1/2$, each of which is described by the Hilbert space \mathcal{H} with $\dim \mathcal{H} = 2$. Initially, the particles interact, resulting in the common spin state described by the vector

$$|\psi\rangle = \frac{1}{\sqrt{2}}[|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle],$$

where the basis vectors

$$|\uparrow\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, |\downarrow\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

describe the states of each particle, with the spin along the positive (resp. negative) direction of the z axis. Usually, one writes briefly

$$|\psi\rangle = \frac{1}{\sqrt{2}}[|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle]. \quad (3.12)$$

Every component denotes the state with the spins in opposite directions, while $|\psi\rangle$ is their superposition, which cannot be represented as a tensor product of state vectors

related to different particles. This state is called “singlet” and is a canonical example of an entangled state of two quantum systems.

Suppose that the particles travel apart along the y axis at a macroscopic distance, while preserving the entangled state of their spins. Consider an experiment in which simultaneous measurements of spin observables $\sigma(\vec{a})$ for one particle and $\sigma(\vec{b})$ for another particle are made simultaneously in remote laboratories A and B . Operators $X = \sigma(\vec{a}) \otimes I, Y = I \otimes \sigma(\vec{b})$ commute. Hence, the corresponding observables are compatible and their covariance is given by expression (2.23).

Exercise 3.14. By using expressions for the matrix elements

$$\langle \uparrow | \sigma(\vec{a}) | \uparrow \rangle = a_z, \quad \langle \downarrow | \sigma(\vec{a}) | \uparrow \rangle = a_x + ia_y, \quad \langle \downarrow | \sigma(\vec{a}) | \downarrow \rangle = -a_z, \quad (3.13)$$

which follows from (2.14), show that in the singlet state the mean value and the variance of the spin observable in any direction are given by

$$E(\sigma(\vec{a}) \otimes I) = 0, \quad D(\sigma(\vec{a}) \otimes I) = 1$$

while the covariance between the spins is given by the formula

$$\langle \psi | \sigma(\vec{a}) \otimes \sigma(\vec{b}) | \psi \rangle = -\vec{a} \cdot \vec{b}. \quad (3.14)$$

It follows that for $\vec{b} = \vec{a}$, the correlation coefficient is equal to -1 . Hence, there is a deterministic relation $a = -b$ between the outcomes a, b of the measurements. The formulas (3.13) and the fact that $\sigma(\vec{a})^2 = I$ imply

$$\langle \psi | [\sigma(\vec{a}) \otimes I + I \otimes \sigma(\vec{a})]^2 | \psi \rangle = 0,$$

whence

$$[\sigma(\vec{a}) \otimes I + I \otimes \sigma(\vec{a})] | \psi \rangle = 0. \quad (3.15)$$

If the spins were classical random variables, this would mean that measuring the spin of the first particle in some arbitrarily chosen direction \vec{a} would “instantaneously” bring the spin of the second particle into the opposite direction.

Exercise 3.15. Prove the following statement: for any direction \vec{a}

$$| \psi \rangle = \frac{\epsilon}{\sqrt{2}} [| \vec{a} \rangle \otimes | -\vec{a} \rangle - | -\vec{a} \rangle \otimes | \vec{a} \rangle], \quad (3.16)$$

where ϵ is inessential factor of modulus one.

Thus, one has to choose between the following alternatives:

1) In quantum mechanics, like in the classical case, the (pure) state describes “real” inner properties of a system. However, in this case, in order to explain how the second

particle “gets to know” the chosen spin direction of the first particle, one has to accept instantaneous action at a distance, contradicting the physical “locality principle”.

2) The state vector is only an expression of the information about the preparation procedure. In that case there is no contradiction with locality, but one has to abandon the complete mechanistic description of a system state as the totality of its “inner properties”.

However, a thorough consideration of the EPR experiment led Bell to a much deeper conclusion: if one tries to describe the quantum correlations classically and simultaneously attempts to satisfy the principle of locality, one can not achieve the level of correlation between systems A and B predicted by quantum theory. “Local realism” and the correlations (3.14) predicted by quantum theory are incompatible.

This is established with the help of the *Clauser–Horne–Shimony–Holt inequality* which is a convenient modification of Bell’s original inequality:

Let X_j, Y_k ($j, k = 1, 2$) be arbitrary random variables on one probability space Ω , such that $|X_j| \leq 1$, $|Y_k| \leq 1$. In this case, for any probability distribution P on Ω , the correlations of these random variables satisfy

$$|EX_1Y_1 + EX_1Y_2 + EX_2Y_1 - EX_2Y_2| \leq 2, \quad (3.17)$$

where E is the expectation corresponding to the distribution P .

The proof is obtained by averaging the elementary inequality over the distribution P

$$-2 \leq X_1Y_1 + X_1Y_2 + X_2Y_1 - X_2Y_2 \leq 2,$$

which in turn follows from

$$|X_1Y_1 + X_1Y_2 + X_2Y_1 - X_2Y_2| \leq |Y_1 + Y_2| + |Y_1 - Y_2| \leq 2 \max\{|Y_1|, |Y_2|\}.$$

Let us go back to the system of two qubits and consider the four different experiments in which the spin observables $\sigma(\vec{a}_j)$, ($j = 1, 2$) are measured over the first qubit, and $\sigma(\vec{b}_k)$, ($k = 1, 2$) in the second, while the directions \vec{a}_j, \vec{b}_k , ($j, k = 1, 2$) will be chosen later. In all four experiments the system is prepared in the singlet state (3.12). Assume that there exists a “local” classical description reproducing the statistical results of all these four experiments. This means that there is a probability space Ω , a probability distribution on Ω describing the statistical ensemble in the singlet state, and the random variables X_j, Y_k ; $j, k = 1, 2$, taking values ± 1 , which classically describe the spin observables $\sigma(\vec{a}_j), \sigma(\vec{b}_k)$; $j, k = 1, 2$. In this case, the quantum correlations (3.14) have to satisfy the inequality (3.17). The requirement of “locality” (or, better to say, *separability* of the two subsystems) is expressed by the fact that the random variables that describe the spin of the first particle (X_1 in the first

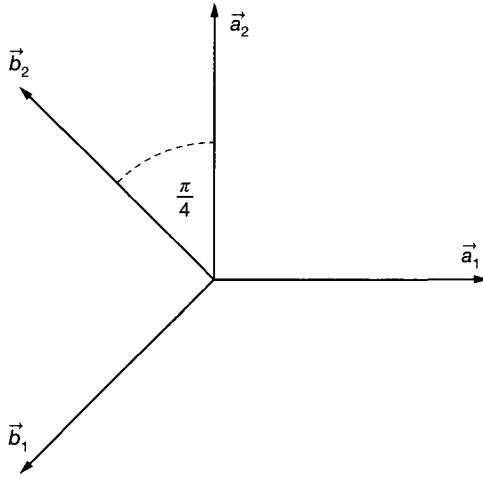


Figure 3.1. Choice of the vectors \vec{a}_j and \vec{b}_k .

two correlations and X_2 in the other two cases) are assumed to be the same for the experiments with different directions of spin of the second particle (Y_1 or Y_2), and *viceversa* for Y_1 and Y_2 , which allows us to apply the inequality (3.17). Omitting this requirement would mean that the given classical description allows the choice of a measurement on the second particle to influence the inner characteristics of the first particle (the term “locality” is used in connection with the fact that the two particles are supposed to be spatially separated).

Now, choose the spin directions as shown in Figure 3.1. Substituting the correlations given by formula (3.14) into the left hand side of (3.17) gives the value $2\sqrt{2}$, which breaks the inequality. Thus, assuming the possibility of a classical “local” description is incorrect. The requirement of separability appears to be so natural that it is not so easy to perceive. However, it is this condition which forbids instantaneous influence of measurements in one subsystem on another. If this condition is abandoned, the four correlations in questions can be arbitrary numbers in $[-1, 1]$, and the left hand side of the inequality (3.17) can be bounded only by the value 4, which does not contradict quantum theory.

Therefore, one has a dilemma: either quantum theory gives incorrect predictions for the correlations, or the composite system of two qubits has no classical local (separable) description. Several experiments were made, starting with Aspect’s experiments in 1981–1982, the results of which essentially confirm the predictions of quantum theory.

3.2.2 Mermin–Peres game

The fact that quantum entanglement between parts of a composite system is profoundly different from any kind of classical connection, and in some situations can provide a real advantage over the latter, is demonstrated in a most spectacular, nearly grotesque, way by “quantum pseudotelepathy games”. We will describe one of these, the *Mermin–Peres magic square game*, which uses nothing but the basic postulates of quantum theory.

In the following game, a team of two players, A (Alice) and B (Bob), tries to win against a croupier C . Imagine a 3×3 -matrix in which C chooses a row and a column and announces the number i of the row to A and the number j of the column – to B . A and B are allowed to agree on any joint strategy prior to this announcement, but cannot communicate after it has been made. Thus, A does not know the number of the column for B and viceversa. Then A must put $+1$ or -1 in each cell of her row, subject to the constraint that the product of elements must be 1 . Similarly B must put $+1$ or -1 in each cell of his row subject to the constraint that the product of elements must be -1 . They win if their choices coincide on the intersection ij of the row and the column chosen by the croupier.

In order to win, players A and B could agree a priori on some fixed 3×3 -matrix with elements ± 1 . However, it is easy to see that there is no matrix that satisfies the imposed constraints: from the constraint on A (resp. B) the product of all matrix elements should be equal to 1 (resp. -1). They could adopt a randomized strategy, but the probability of success will always be strictly less than 1 .

However, assuming that they are capable of producing and sharing an entangled quantum state before C makes his announcement, and make a pre-agreed quantum measurements on their respective parts of this state after the announcement, they will always be able to win! Of course, both the entangled state and the measurements are independent of the choices of C , as it should be. Hence, the rules of the game are perfectly respected and it is the use of advanced quantum information technology that produces a result that is inachievable by solely classical means. From the point of view of a purely classical observer it looks like there is some immaterial connection between A and B , hence the name “pseudotelepathy”.

For a pair of qubits A and B , let us introduce the *Bell state*

$$|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle)$$

which is similar to the singlet state (3.12).

The prepared entangled state is the tensor product of two Bell states of two pairs of qubits $A_1 B_1$ and $A_2 B_2$:

$$|\Psi\rangle = |\Psi\rangle_{A_1 B_1} \otimes |\Psi\rangle_{A_2 B_2}.$$

Before C makes his announcement, qubits A_1 and A_2 are distributed to player A , while qubits B_1 and B_2 are given to B . Players A and B also agree in advance that,

as soon as they receive the numbers of the row i and the column j , respectively, they will perform the spin measurements on their pairs of qubits, according to the following table

$$\begin{bmatrix} \sigma_0 \otimes \sigma_z & \sigma_z \otimes \sigma_0 & \sigma_z \otimes \sigma_z \\ \sigma_x \otimes \sigma_0 & \sigma_0 \otimes \sigma_x & \sigma_x \otimes \sigma_x \\ -\sigma_x \otimes \sigma_z & -\sigma_z \otimes \sigma_x & \sigma_y \otimes \sigma_y \end{bmatrix}$$

and write the outcome of the measurement in the corresponding cell. At this point, we should expand on the special properties of this array of operators.

Denoting by X_{ij} the operator on the intersection of i -th row and j -th column, one has

- i. $X_{ij} = X_{ij}^*$ and $X_{ij}^2 = \sigma_0 \otimes \sigma_0 \equiv I$, so that all X_{ij} are (sharp) real observables with eigenvalues ± 1 ;
- ii. in any row i , the operators X_{ij} ; $j = 1, 2, 3$ commute. Hence, the observables are compatible. Moreover, $X_{i1}X_{i2}X_{i3} = I$. Hence for any value $i = 1, 2, 3$ given by C player A can make a joint measurement of the observables X_{ij} ; $j = 1, 2, 3$, obtaining the outcomes $+1$ or -1 , which obey the constraint for A . Next, A writes these outcomes in row i . A similar description applies to player B and any column j .
- iii. miraculously, the numbers written on the intersection of i -th row and j -th column by A and B always coincide. This follows from the identity

$$(X_{ij}^A \otimes X_{ij}^B)|\Psi\rangle = |\Psi\rangle; \quad i, j = 1, 2, 3,$$

where X_{ij}^A (resp. X_{ij}^B) is the operator X_{ij} in system $A = A_1A_2$ (resp. $B = B_1B_2$). Indeed, this identity means that, if the whole system $A_1A_2B_1B_2$ is prepared in the state $|\Psi\rangle$, the product of the outcomes of the (compatible) measurements of A and B , in any cell ij , will be equal to 1, i.e. the outcomes coincide.

Exercise 3.16. Prove the properties i., ii. by using the multiplication table (2.17) for the Pauli matrices. Prove the property iii. by using the identities

$$\begin{aligned} (\sigma_{x,z} \otimes \sigma_0)|\Psi\rangle_{AB} &= (\sigma_0 \otimes \sigma_{x,z})|\Psi\rangle_{AB}; \\ (\sigma_y \otimes \sigma_0)|\Psi\rangle_{AB} &= -(\sigma_0 \otimes \sigma_y)|\Psi\rangle_{AB} \end{aligned}$$

similar to (2.17).

3.3 Quantum systems as information carriers

3.3.1 Transmission of classical information

If the information carrier is a classical system with a finite number of states d , the maximal number of binary digits, or *bits*, which can be recorded or transmitted by such a carrier is equal to $\log_2 d$. If the transmission is without errors, one speaks of an ideal communication channel. In general, a noisy classical channel is described by the conditional probabilities $p(y|x)$ for receiving the signal y at the output, if the value at the input was x . For the ideal channel $p(y|x) = \delta_{xy}$. The opposite case is the channel for which the probabilities $p(y|x) = p(y)$ do not depend on the transmitted signal x and hence the information is not transmitted at all. More detailed account of classical channels will be given in Ch. 4.

Let us now consider how classical information can be transmitted with a quantum carrier described by a Hilbert space \mathcal{H} . The carrier's state S is prepared with macroscopic devices. By changing the parameters of a device, one changes the parameters of the state, thus having a possibility to “record” classical information into the quantum state. Let there be n different signals and S_x be the quantum state corresponding to the signal $x = 1, \dots, n$. The mapping $x \rightarrow S_x$ describes the net result of a physical process that generates the state S_x . A detailed description of such a process is beyond the scope of information theory, for which only the resulting states S_x are relevant.

To extract the classical information contained in the output state of a channel, one has to perform a measurement. If one measures the observable $M = \{M_y\}$, the conditional probability of obtaining a signal y at the output given the input signal x is equal to

$$p_M(y|x) = \text{Tr } S_x M_y. \quad (3.18)$$

Thus, for a fixed quantum measurement, we have the usual classical channel. Assume that one manages to prepare states and perform a measurement in such a way that $S_x = |e_x\rangle\langle e_x|$, $M_y = |e_y\rangle\langle e_y|$, then $P_M(y|x) = \delta_{xy}$, i.e. one has an ideal classical channel, capable of transmitting $\log_2 d$ bits, where $d = \dim \mathcal{H}$. Later, in Chapter 5, we will establish that this quantity is the upper bound for the amount of classical information that can in principle be transmitted with the given quantum carrier. This implies the following important consequences:

- 1) The fact that Hilbert space contains infinitely many different state vectors does not aid us in transmitting unlimited amount of information. The more states are used for transmission, the closer they are to each other and hence they become less and less distinguishable.
- 2) The dimensionality of the Hilbert space is a measure of the ultimate information resource of the corresponding quantum system.

3.3.2 Entanglement and local operations

In this section, we will need elementary knowledge about quantum evolutions. Later, in Ch. 6, we will provide a detailed discussion of this question from the viewpoint of the general theory. Meanwhile, it is sufficient to know that reversible evolutions of a quantum system are described by unitary operators U . The vector of the initial pure state ψ is transformed into $U\psi$. Correspondingly, a density operator S is transformed into USU^* .

Consider now the following question. The nonlocal, from a classical viewpoint, nature of the EPR-correlations suggests to attempt using them for instantaneous information transmission. Let us show that this is not possible within the quantum mechanical framework, so that the EPR correlations do not contradict locality. Consider two quantum systems A and B , in the corresponding spaces \mathcal{H}_A and \mathcal{H}_B , which are in an entangled state S_{AB} . In the case of practical interest, the systems are sufficiently spatially separated, although this is not reflected by the formulas. The system A (transmitter) uses the classical messages x , which should be transmitted to B (receiver), in order to perform certain unitary operations U_x in the space \mathcal{H}_A (such operations are called *local*, i.e. acting nontrivially only in one of the two systems). This brings the state of the system AB into $S_x = (U_x \otimes I_B)S_{AB}(U_x \otimes I_B)^*$, and in this way the classical information is written into the quantum state of the composite system. In this case, observer B makes a local measurement of an observable in \mathcal{H}_B , corresponding to the resolution of the identity $M = \{I_A \otimes M_y\}$ in the space of the composite system $\mathcal{H}_A \otimes \mathcal{H}_B$. The resulting conditional probability (3.18) is

$$p_M(y|x) = \text{Tr} (U_x \otimes I_B)S_{AB}(U_x \otimes I_B)^*(I_A \otimes M_y) = \text{Tr} S_B M_y,$$

where $S_B = \text{Tr}_A S_{AB}$ is the partial state of the system B , which does not depend on x . This means that local operations of observer A do not in any way influence the state of B , and hence, no information is transmitted.

Consider now the different scenario where observer A makes a local measurement of observable M_x^A and sends the outcome to observer B , who takes this into account by selecting the subensemble where this outcome x has appeared. In this case, the *posterior state* $S_B(x)$ of the party B will depend on the outcome of the measurement x . To find the posterior state, assume that B performs his local measurement M_y^B , and consider the joint probability distribution of the two measurements

$$p_{x,y} = \text{Tr} S_{AB}(M_x^A \otimes M_y^B).$$

By introducing the conditional probabilities $p(y|x) = p_{x,y}/p_x$ (assuming $p_x > 0$), we have

$$\begin{aligned} p_x &= \text{Tr} S_{AB}(M_x^A \otimes I_B) = \text{Tr} S_A M_x^A, \\ p(y|x) &= \text{Tr} S_B(x) M_y^B, \end{aligned}$$

where

$$S_B(x) = p_x^{-1} \text{Tr } \mathcal{H}_A S_{AB} (M_x^A \otimes I_B)$$

are the posterior states.

The change of the state of observer B from S_B to $S_B(x)$ means a reduction of the full statistical ensemble of B to the subensemble characterized by the value x of the outcome of the measurement of observer A . If this outcome is not taken into account (for example, A does not send it to B , hence B makes no selection), then the state of B is not changed

$$S_B = \sum_x p_x S_B(x).$$

Exercise 3.17. Let $S_{AB} = |\psi\rangle\langle\psi|$ be a pure entangled state,

$$|\psi\rangle = \sum_x |e_x^A\rangle \otimes |\psi_x^B\rangle,$$

where e_x^A is an orthonormal basis in \mathcal{H}_A , $\sum_x \|\psi_x^B\|^2 = 1$. Assume that A measures the observable $M_x^A = |e_x^A\rangle\langle e_x^A|$ and sends the measurement outcome x to the party B . Prove that $p_x = \|\psi_x^B\|^2$ and the posterior states of B are

$$S_B(x) = p_x^{-1} |\psi_x^B\rangle\langle\psi_x^B|. \quad (3.19)$$

Coming back to the EPR experiment, we obtain from the relation (3.16), that in the case where A measures the spin observable $\sigma(\vec{a})$ and sends the outcome ± 1 to observer B , the posterior state of B (i.e. the state of the corresponding statistical subensemble) is $|\mp \vec{a}\rangle\langle \mp \vec{a}|$. On the other hand, if there is no communication between A and B , the state of B remains chaotic, independently of the measurement of A . Thus, quantum theory does not contradict the principle of locality.

3.3.3 Superdense coding

Although EPR-correlations cannot aid in transmitting information from A to B , it appears that using these correlations as additional resource allows one to double the amount of classical information transmitted from A to B if they are connected by an ideal quantum channel enabling perfect transmission of a quantum state from A to B . Hence, EPR correlations appear as a “catalyst” for classical information transmission and from this viewpoint can be considered a unit of a specific information resource (sometime called an *ebit*).

Consider the two systems A and B , each of which is a qubit, connected by an ideal quantum channel. The maximum amount of classical information that can be

transmitted from A to B is equal to 1 bit, and is obtained by encoding this bit into two orthogonal vectors, say,

$$0 \rightarrow |0\rangle, \quad 1 \rightarrow |1\rangle,$$

where $|0\rangle, |1\rangle$ is the canonical basis for one qubit.

The protocol of “superdense coding”, which allows us to double the amount of transmitted classical information, is based on the following simple mathematical fact: all vectors of the *Bell basis*

$$|e_+\rangle = |00\rangle + |11\rangle, \quad |e_-\rangle = |00\rangle - |11\rangle, \quad |h_+\rangle = |10\rangle + |01\rangle, \quad |h_-\rangle = |10\rangle - |01\rangle$$

in the system of two qubits AB (we will systematically omit the normalizing factor $1/\sqrt{2}$) can be obtained from one of these vectors by the action of “local” unitary operators, i.e. operators acting nontrivially only on qubit A . Namely,

$$|e_-\rangle = (\sigma_z \otimes I)|e_+\rangle, \quad |h_+\rangle = (\sigma_x \otimes I)|e_+\rangle, \quad |h_-\rangle = -i(\sigma_y \otimes I)|e_+\rangle.$$

Thus, if AB is initially in the state $|e_+\rangle$, A can encode 2 bits of classical information into 4 states of the Bell basis by mere local operations $\sigma_\gamma; \gamma = 0, x, y, z$, and then (physically) send its qubit to B via the ideal quantum channel. Then, making the measurement in the Bell basis, B receives 2 bits of classical information.

3.3.4 Quantum teleportation

So far, we considered transmission of classical information. This information can be “written” into a quantum state and transmitted through a physical channel. However, the quantum state itself is an information resource, since it carries statistical uncertainty. The information carried by an unknown quantum state is qualitatively different from the classical one, and deserves the special name *quantum information*. The most apparent distinction of quantum information is impossibility of copying (*no cloning*). While the classical information can be reproduced in an arbitrary number of copies, there is no room for a physical device that could perform the same task for quantum information. Indeed, the cloning transformation

$$|\psi\rangle \rightarrow \underbrace{|\psi\rangle \otimes \cdots \otimes |\psi\rangle}_n$$

is nonlinear, and hence cannot be implemented by a unitary operator. Of course, any fixed state (or even a collection of orthogonal states) may be copied by a specific device, but there is no universal device which would do this for an arbitrary state.

Let us briefly discuss the question of how a quantum state can be transmitted. A straightforward way is to just send the quantum system itself. Much more interesting and nontrivial is *teleportation* of the state, where the system itself is not transmitted, but instead a certain classical message is sent. An essential additional resource

are the EPR-correlations between the input and the output, which again play the role of “catalyst”. Note that teleportation of a quantum state cannot be accomplished by merely sending classical information without using the entanglement. Since the classical information can be copied, this would mean the possibility of cloning quantum information.

Let there be two quantum systems A and B , describing input and output of the communication channel, respectively. In the simplest and basic version, the systems A and B are just qubits.

- i. Before the transmission, the system AB is prepared in the state $|00\rangle + |11\rangle$. (Recall that we omit the normalizing factor $1/\sqrt{2}$).
- ii. The third party C provides A with an arbitrary pure state

$$|\psi\rangle = a|0\rangle + b|1\rangle,$$

which should be transmitted to B . Now, the combination of the three systems CAB is in the state

$$(a|0\rangle + b|1\rangle) \otimes (|00\rangle + |11\rangle).$$

- iii. Then observer A performs a certain reversible local transformation of the state of CA
- iv. Observer A performs a measurement in CA with 4 outcomes (comprising 2 bits of classical information) and transmits the outcome of its measurement to B by using the ideal classical communication channel. The transformation and the measurement will be described below.
- v. Depending on the outcome received, observer B performs a certain transformation of its state, obtaining the required state vector $|\psi\rangle$.

The necessary transformations are examples of quantum “gates”, used in quantum computations. In step iii. the state of CA is transformed by the operation CNOT (controlled “not”):

$$|00\rangle \rightarrow |00\rangle, \quad |01\rangle \rightarrow |01\rangle, \quad |10\rangle \rightarrow |11\rangle, \quad |11\rangle \rightarrow |10\rangle,$$

when the state of the first qubit is unchanged, whilst the state of the second qubit is changed to its opposite if and only if the state of the first qubit is $|1\rangle$. As this transformation turns one basis into another basis, it is a unitary operator in the 4-dimensional space of CA . Now, the qubit C is transformed by the *Hadamard gate* H , implemented by the unitary matrix

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Then the basis is rotated by the angle $\pi/4$:

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |1\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle).$$

The initial state of the total system CAB is

$$a|000\rangle + b|100\rangle + a|011\rangle + b|111\rangle.$$

The action of CNOT on CA results in

$$a|000\rangle + b|110\rangle + a|011\rangle + b|101\rangle.$$

Then the Hadamard gate H acts on C , and we obtain

$$a(|000\rangle + |100\rangle) + b(|010\rangle - |110\rangle) + a(|011\rangle + |111\rangle) + b(|001\rangle - |101\rangle).$$

Singling out the state of the subsystem CA we have

$$|00\rangle(a|0\rangle + b|1\rangle) + |01\rangle(a|1\rangle + b|0\rangle) + |10\rangle(a|0\rangle - b|1\rangle) + |11\rangle(a|1\rangle - b|0\rangle).$$

Now, in step iv., the measurement in the subsystem CA is performed, which projects onto one of four basis vectors $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. According to formula (3.19), the posterior state of the system B , depending on the obtained measurement outcome, is described by one of the vectors

$$a|0\rangle + b|1\rangle, \quad a|1\rangle + b|0\rangle, \quad a|0\rangle - b|1\rangle, \quad a|1\rangle - b|0\rangle.$$

The outcome of the measurement is sent to B through the ideal classical channel. In step v, depending on the outcome, B applies one of the unitary Pauli operators $\sigma_0 = I, \sigma_x, \sigma_z, \sigma_y$, in each case transforming the state of B into $a|0\rangle + b|1\rangle$. Thus, observer B is transformed to the state in which observer C was initially, while the state of C is irreversibly lost (otherwise cloning of the quantum information would be possible).

3.4 Notes and references

1. For more detail about tensor product of Hilbert spaces, see Ch. II.4 of the book [171], where one can also find solutions to the Exercises in this section. For a general formulation of Naimark's Dilation Theorem, see [159], where it was proved for an arbitrary resolution of the identity (probability operator-valued measure, see Definition 11.29) in infinite dimensional Hilbert space. The motivation was to obtain self-adjoint extensions of symmetric operators but it was later recognized that this was just the first discovery of the remarkable general phenomenon that the properties of various operator objects can be improved by extension to a wider underlying space. In

Quantum Information Theory this principle received the colorful name “the Church of Larger Hilbert Space”. The purification of quantum states and the unitary dilation of irreversible quantum dynamics in Section 6.3 are two of its other manifestations.

An analog of the decomposition (3.8) in separable Hilbert space dates back to Schmidt’s paper of 1907, which is devoted to integral equations. The “purification map” was introduced by Powers and Störmer [169] and studied by Woronowicz [229] in connection with the problem of the quasi-equivalence of states on a C^* -algebra. The trick of purifying mixed states, which is widely used in quantum information theory, also appeared in the paper of Lindblad [149].

The observation that the PPT condition is necessary for separability is due Peres [166] and Horodeckis [115], while the latter had shown that this condition is also sufficient for separability of states in 2×2 - and 2×3 -dimensional composite systems. In higher dimensions there exist PPT states that are not separable, but their entanglement (called *bound entanglement*) is “weak”, since it doesn’t allow for *distillation*, i.e. transformation to maximally entangled Bell states by using local operations and classical communication (LOCC) (P. Horodecki [116], Horodeckis [119]).

2. The term “entanglement” (German “Verschränktheit”) was introduced by Schrödinger, who was the first to notice the unusual properties of entangled states. The careful logical analysis of the EPR experiment is due to Bell [18], who in particular established an inequality relative to (3.17) and applied it to analyze the EPR paradox. In the discussion of the Mermin–Peres game, we followed Aravind [8], where one can find further references. Experimental demonstrations of the quantum gain were also reported, see Zu et al. [231].

3. The protocol of superdense coding was proposed by Bennett and Wiesner [26]. The protocol of quantum teleportation was proposed in the famous publication of Bennett, Brassard, Crépeau, Jozsa, Peres and Wootters [19]. The possibility of teleportation of the polarization state of a photon was experimentally demonstrated by the group of Zeilinger in 1997. Since then, a number of further successful experiments were performed, both on photons and on atoms.

Part II

The primary coding theorems

Chapter 4

Classical entropy and information

4.1 Entropy of a random variable and data compression

Let X be a random variable with values in a finite set \mathcal{X} and let $P = \{p_x; x \in \mathcal{X}\}$ be its probability distribution. The *entropy* $H(X) = H(P)$ is defined as

$$H(X) = \sum_{x \in \mathcal{X}} \eta(p_x), \quad (4.1)$$

where

$$\eta(t) = \begin{cases} -t \log t, & t > 0, \\ 0, & t = 0, \end{cases} \quad (4.2)$$

Here and in what follows $\log = \log_2$ denotes the binary logarithm. The entropy $H(X)$ is a measure of uncertainty, of variability, or of the information content in the random variable X , as we shall see below. Without loss of generality we will assume in what follows that $p_x > 0$ for all x .

Let d be the number of elements in the set \mathcal{X} . One has

$$0 \leq H(P) \leq \log d, \quad (4.3)$$

with the minimum value attained on degenerate distributions $\{1, 0, \dots, 0\}, \dots, \{0, \dots, 0, 1\}$, and the maximum value on the uniform distribution $\{\frac{1}{d}, \dots, \frac{1}{d}\}$. The first inequality follows from the non-negativity of the function $\eta(t)$ in the domain $[0, 1]$, and the second from concavity of the function $t \rightarrow \log t$:

$$H(P) = \sum_{x=1}^d p_x \log \frac{1}{p_x} \leq \log \sum_{x=1}^d p_x \frac{1}{p_x} = \log d.$$

Let us now explain the operational interpretation of the entropy $H(X)$ as a measure of the information content in the random variable X . Consider a “random source” which produces independent, identically distributed (i.i.d.) random variables with distribution P . A sequence $w = (x_1, \dots, x_n)$ of n letters from the alphabet \mathcal{X} is called a *word* of length n . The total number of such words is $|\mathcal{X}|^n = 2^{n \log |\mathcal{X}|}$. Hence it is possible to encode all these words using binary strings of length $n \log |\mathcal{X}|$, i.e. $n \log |\mathcal{X}|$ bits¹. However, by using the fact that the probability distribution P of X

¹ To avoid inessential complications, we shall systematically neglect the fact that this number is not necessarily an integer.

is in general non-uniform, one can in general achieve a much better encoding. The possibility of the data compression is closely connected with the *Asymptotic Equipartition Property*.

Theorem 4.1. *If X_1, \dots, X_n are i.i.d. random variables with distribution P , then*

$$-\frac{1}{n} \sum_{i=1}^n \log p_{X_i} \rightarrow H(X) \quad \text{in probability.} \quad (4.4)$$

This means that for any $\delta, \epsilon > 0$ there exists an n_0 , such that for all $n \geq n_0$ we have

$$P \left\{ \left| -\frac{1}{n} \sum_{i=1}^n \log p_{X_i} - H(X) \right| < \delta \right\} > 1 - \epsilon. \quad (4.5)$$

This is a direct consequence of the *Law of Large Numbers* applied to the i.i.d. random variables $-\log p_{X_i}; i = 1, \dots, n$, taking into account that $E(-\log p_{X_i}) = H(X)$.

Noting that the probability of the occurrence of a word $w = (x_1, \dots, x_n)$ is

$$p_w = p_{x_1} \cdots p_{x_n} = 2^{-n \left(-\frac{1}{n} \sum_{i=1}^n \log p_{x_i} \right)} \quad (4.6)$$

we can now use relation (4.5) to introduce a basic notion of the *typical word*:

Definition 4.2. A word w , having probability p_w , is called δ -typical if

$$2^{-n(H(X)+\delta)} < p_w < 2^{-n(H(X)-\delta)}, \quad (4.7)$$

in other words, if it satisfies the event in the formula (4.5).

The collection of all δ -typical words of length n will be denoted by $T^{n,\delta}$. By using (4.5), (4.7), it is straightforward to see that δ -typical words have the following properties (where δ, ϵ are fixed positive numbers):

Exercise 4.3. Prove the following statements:

- i. There are at most $2^{n(H(X)+\delta)}$ typical words, i.e. $|T^{n,\delta}| \leq 2^{n(H(X)+\delta)}$.
- ii. For sufficiently large n , the set of all non-typical words has a probability $P(\overline{T^{n,\delta}}) \leq \epsilon$.
- iii. For sufficiently large n , there are at least $(1 - \epsilon)2^{n(H(X)-\delta)}$ typical words.

Now, one can do efficient *data compression* by using all binary sequences of length $n(H(X) + \delta)$ to encode all δ -typical words, and all non-typical words into one additional (idle) symbol. In this case, the error probability of such a coding scheme will be less than or equal to ϵ .

Conversely, any code that uses only binary sequences of length $n(H(X) - \delta)$ has an asymptotically non-vanishing error probability, which in fact can be made arbitrarily close to 1 for n large enough. Indeed, let C be the collection of words that are used for the encoding into $n(H(X) - \delta)$ binary strings, while all other words are encoded into the idle symbol. In this case, the error probability is $1 - P(C)$, where

$$P(C) = P(C \cap T^{n, \delta/2}) + P(C \cap \overline{T^{n, \delta/2}}) \quad (4.8)$$

$$\leq |C|2^{-n(H(X) - \delta/2)} + P(\overline{T^{n, \delta/2}}) \quad (4.9)$$

$$\leq 2^{-n\delta/2} + \epsilon, \quad (4.10)$$

which can be made arbitrarily small for n large enough.

Since we need asymptotically $N \sim 2^{nH(X)}$ codewords for efficient encoding, $H(X)$ can be interpreted as a measure of the information content of the source in bits per letter. It is also clear that if we start with a uniform distribution $p_x = \frac{1}{|\mathcal{X}|}$, $H(X) = \log |\mathcal{X}|$ and no data compression is possible.

Exercise 4.4. Let \mathcal{X} be a finite subset of \mathbb{R} and let X_1, \dots, X_n, \dots be i.i.d. random variables with values in \mathcal{X} . Assuming the conditions of Theorem 4.1, prove the following large deviations inequality:

$$P\left\{\frac{1}{n} \sum_{i=1}^n (X_i - EX) \geq \delta\right\} \leq \exp\left\{-n \sup_{s>0} [s\delta - \mu(s)]\right\}, \quad (4.11)$$

where

$$\mu(s) = \ln Ee^{s(X-EX)} = o(s) \quad \text{for } s \rightarrow 0.$$

This implies the Law of Large Numbers with exponential decay of tails, since $s\delta - \mu(s) > 0$ for small $s > 0$.

Hint: prove and apply the following version of the Markov inequality

$$P\{X \geq x\} = P\{e^{sX} \geq e^{sx}\} \leq \exp(-sx) Ee^{sX}$$

for any $x \in \mathbb{R}$ and $s > 0$.

4.2 Conditional entropy and the Shannon information

Let X, Y be random variables on one probability space with joint probability distribution $\{p_{x,y}\}$. In this case, one can define their *joint entropy* $H(XY)$ similar to

relation (4.1). The *conditional entropy* $H(Y|X)$ is defined as

$$\begin{aligned}
 H(Y|X) &= \sum_x p_x H(Y|X = x) \\
 &= - \sum_x p_x \sum_y p(y|x) \log p(y|x) \\
 &= - \sum_{x,y} p_{x,y} \log p_{x,y} + \sum_x p_x \log p_x \\
 &= H(XY) - H(X).
 \end{aligned} \tag{4.12}$$

There is a useful general rule stating that any linear relation for (conditional) entropies continues to hold if, in every term, a condition is supplemented with one and the same extra condition. For example, (4.12) implies

$$H(XY|Z) = H(X|Z) + H(Y|XZ). \tag{4.13}$$

A noisy *communication channel* is described by the transition probability $p(y|x)$ from the input alphabet \mathcal{X} to the output alphabet \mathcal{Y} , i.e. a collection of conditional probabilities that a letter $y \in \mathcal{Y}$ is received when a letter $x \in \mathcal{X}$ is sent. An input probability distribution $P = \{p_x\}$ is transformed by the channel to the output probability distribution $P' = \{p'_y\}$, where $p'_y = \sum_x p(y|x)p_x$. The *Shannon information* (about the random variable X contained in Y) is defined as

$$I(X; Y) = H(Y) - H(Y|X), \tag{4.14}$$

where the entropy $H(Y)$ can be interpreted as the information content of the output, while the conditional entropy $H(Y|X)$ is interpreted as its useless component that arises from the *noise* in the channel. Substituting (4.12) into this expression for the Shannon information, we see that it is symmetric in X and Y , and hence it can also be called the *mutual information*

$$I(X; Y) = H(X) + H(Y) - H(XY). \tag{4.15}$$

Furthermore, $I(X; Y) = H(X) - H(X|Y)$, where $H(X)$ can now be interpreted as the information content of the input, and $H(X|Y)$ is sometimes referred to as *loss*. An explicit expression for the Shannon information via the input distribution and the channel transition probabilities is given by

$$I(X; Y) = \sum_{xy} p_x p(y|x) \log \left(\frac{p(y|x)}{\sum_{x'} p(y|x') p_{x'}} \right). \tag{4.16}$$

A very useful quantity from classical statistics is the *relative entropy* of two probability distributions $P = \{p_x\}$, $Q = \{q_x\}$ (the Kullback–Leibler–Sanov entropy):

$$H(P; Q) = \begin{cases} \sum_{x: p_x > 0} p_x \log \frac{p_x}{q_x}, & \text{if } \{x : p_x > 0\} \subseteq \{x : q_x > 0\}; \\ +\infty & \text{otherwise.} \end{cases}$$

By using the inequality $\ln t \leq t - 1$, we obtain

$$H(P; Q) \geq -\log e \sum_{x: p_x > 0} p_x \left(\frac{q_x}{p_x} - 1 \right) \geq 0, \quad (4.17)$$

with equality if and only if $P = Q$.

The quantity $H(P; Q)$ plays an important role as an “asymmetric distance” between the probability distributions. It has the following monotonicity property: consider the channel $r(y|x)$ and two input probability distributions $P = \{p_x\}$, $Q = \{q_x\}$, which are transformed to the output distributions $P' = \{p'_y\}$, $Q' = \{q'_y\}$,

$$p'_y = \sum_x r(y|x)p_x, \quad q'_y = \sum_x r(y|x)q_x.$$

In this case,

$$H(P'; Q') \leq H(P; Q). \quad (4.18)$$

In fact,

$$\begin{aligned} H(P; Q) &= \sum_{xy} p_x r(y|x) \log \frac{r(y|x)p_x}{r(y|x)q_x} \\ &= \sum_{xy} p_{x,y} \left(\log \frac{p'_y}{q'_y} + \log \frac{p(x|y)}{q(x|y)} \right) \\ &= H(P'; Q') + \sum_y p'_y H(p(x|y); q(x|y)) \\ &\geq H(P'; Q'). \end{aligned}$$

Exercise 4.5. Prove that $I(X; Y) = H(p_{x,y}; p_x \cdot p_y)$. Hence, due to (4.17), $I(X; Y) \geq 0$, while $I(X; Y) = 0$ if and only if X and Y are independent random variables: $p_{x,y} = p_x \cdot p_y$.

This also implies monotonicity of the conditional entropy

$$0 \leq H(Y|X) \leq H(Y),$$

and sub-additivity of the entropy,

$$H(XY) \leq H(X) + H(Y). \quad (4.19)$$

Exercise 4.6. Consider the concatenation of two channels $p(y|x)$ and $q(z|y)$, defined as $r(z|x) = \sum_y q(z|y)p(y|x)$. Let X be a random variable at the input of the channel $p(y|x)$ having the probability distribution $\{p(x)\}$, and denote by Y and Z , the output of the channel $p(y|x)$ and $r(z|x)$, resp., so that $X \rightarrow Y \rightarrow Z$ is a Markov chain. In this case, the *data-processing inequalities* hold,

$$I(X; Z) \leq \min\{I(X; Y), I(Y; Z)\}. \quad (4.20)$$

Hint: $I(X; Z) \leq I(X; YZ) = I(X; Y)$ follows from $H(X|Z) \geq H(X|YZ) = H(X|Y)$ and analogously for the second quantity.

The Shannon information has the following expression, in terms of the relative entropy

$$I(X; Y) = \sum_x p_x H(P'_x; P'), \quad (4.21)$$

where $P'_x = \{p(y|x)\}$ for fixed $x \in \mathcal{X}$, and $P' = \sum_x p_x P'_x$ is the distribution of Y .

4.3 The Shannon capacity of the classical noisy channel

Consider the classical channel $X \rightarrow Y$ determined by the transition probability $p(y|x)$. The most important characteristic of the channel is the *Shannon capacity*

$$C_{\text{Shan}} = \max_X I(X; Y), \quad (4.22)$$

where the maximum is taken over all possible distributions $P = \{p_x\}$ of the input X . It will later be identified as the operationally defined information transmission capacity of the channel.

As an example, let us consider the *binary symmetric channel*. Assume that \mathcal{X} and \mathcal{Y} consist of two letters 0, 1, which are transmitted with probability p and flipped with probability $1 - p$. Introducing the *binary entropy*

$$h_2(p) = -p \log p - (1 - p) \log(1 - p), \quad (4.23)$$

the mutual information can be written as $I(X; Y) = H(Y) - H(Y|X) = H(Y) - h_2(p)$, and is thus bounded by the quantity

$$C_{\text{Shan}} = 1 - h_2(p), \quad (4.24)$$

which is achieved for the uniform input distribution: $p_0 = p_1 = \frac{1}{2}$.

In general, $I(X; Y)$ is a concave function of the input probability distribution $P = \{p_x\}$. Hence, the Kuhn–Tucker conditions from convex analysis, see e.g. [42], can be applied to characterize the optimal input distribution P that maximizes $I(X; Y)$.

Exercise 4.7. Show that the entropy $H(P)$ is a concave function of the probability distribution P . Hint: use concavity of the function $t \rightarrow \eta(t)$. Therefore, the output entropy $H(Y) = H(P')$ is also concave, while the conditional entropy $H(Y|X)$ is affine. Use formula (4.14) to conclude that $I(X; Y)$ is concave.

Exercise 4.8. (*Kuhn–Tucker conditions*) Let $F(P)$ be a concave function of probability distribution P , continuously differentiable at the interior of the simplex of all probability distributions on \mathcal{X} , and having possibly infinite limits of the partial derivatives at its boundary. Prove the following statement: a necessary and sufficient condition for the point P^0 to be the maximizer for $F(P)$ is that there exists a λ such that

$$\frac{\partial F(P^0)}{\partial p_x} \begin{cases} = \lambda, & \text{if } p_x^0 > 0; \\ \leq \lambda, & \text{if } p_x^0 = 0. \end{cases}$$

In the case of Shannon information (4.21), we find

$$\frac{\partial I}{\partial p_x} = H(P'_x; P') - \log e,$$

which leads to the following “maximal distance” characterization of the optimal input distribution $P^0 = \{p_x^0\}$: there exists a μ such that

$$H(P'_x; (P^0)') \begin{cases} = \mu, & \text{if } p_x^0 > 0; \\ \leq \mu, & \text{if } p_x^0 = 0, \end{cases} \quad (4.25)$$

in which case, by necessity $\mu = C_{\text{Shan}}$.

Assume that two channels $\{p_j(y_j|x_j)\}; j = 1, 2$, are given, and consider the composite parallel channel consisting of the independent use of these two channels, described by the transition probability $\{p_1(y_1|x_1)p_2(y_2|x_2)\}$.

Proposition 4.9. *The Shannon capacity of the composite channel is*

$$(C_{\text{Shan}})_{12} = (C_{\text{Shan}})_1 + (C_{\text{Shan}})_2. \quad (4.26)$$

Thus the Shannon capacity is additive for the parallel channels.

Proof. Let P^j be the optimal input probability distributions for the j -th channel. In this case, they satisfy the condition (4.25) with constants $\mu_j = (C_{\text{Shan}})_j$. Using the fact that in general

$$H(P^1 \times P^2; Q^1 \times Q^2) = H(P^1; Q^1) + H(P^2; Q^2),$$

we obtain that the probability distribution $P = P^1 \times P^2$ satisfies the condition (4.25) for the composite channel, with the constant $\mu = (C_{\text{Shan}})_1 + (C_{\text{Shan}})_2$. \square

4.4 The channel coding theorem

Given the channel $p(y|x)$, one can consider the composite *memoryless* channel

$$p(y^n|x^n) = p(y_1|x_1) \cdots p(y_n|x_n), \quad (4.27)$$

which transmits words of length n , by using the initial channel n times, with each use being independent of the others,

$$x^n \left\{ \begin{array}{ccc} x_1 & \longrightarrow & y_1 \\ x_2 & \longrightarrow & y_2 \\ \vdots & & \vdots \\ x_n & \longrightarrow & y_n \end{array} \right\} y^n,$$

By X^n, Y^n , we denote the random variables at the input and the output of the composite channel.

Following (4.22), consider the quantity

$$C_n = \max_{X^n} I(X^n; Y^n)$$

for the composite channel. Inductively applying property (4.26), we find that the sequence $\{C_n\}$ is *additive* for memoryless channels, whence

$$C_n = nC_{\text{Shan}}. \quad (4.28)$$

To reduce the effect of noise, one uses encoding of the messages at the input and, correspondingly, decoding at the output. The whole process of information transmission is given by the diagram:

$$i \xrightarrow{\text{encoding}} x^n \xrightarrow{\text{channel}} y^n \xrightarrow{\text{decoding}} j, \quad (4.29)$$

where i (resp. j) denote transmitted (resp. received) messages. One can simply assume that $i, j \in \{1, \dots, N\}$ are the indices of the messages, since only the quantity N of the transmitted messages is of importance.

The aim is to choose an encoding and decoding that maximizes the *transmission rate* $R = \frac{\log N}{n}$, equal to number of bits per transmitted symbol, while keeping the error probability small. Let us proceed to the exact formulation.

Definition 4.10. A *code* (W, V) of size N for the classical memoryless channel consists of a codebook W , which is a collection of N codewords $w^{(1)}, \dots, w^{(N)}$ of length n , and a decision rule that is a decomposition of \mathcal{Y}^n into $N + 1$ nonintersecting subsets $V = \{V^{(0)}, V^{(1)}, \dots, V^{(N)}\}$. The subsets $V^{(1)}, \dots, V^{(N)}$ can be interpreted as decision domains. Whenever $y^n \in V^{(j)}$ is received, we decide that the word $w^{(j)}$ was transmitted. If $y^n \in V^{(0)} = \overline{\cup_{j=1}^N V^{(j)}}$, no definite decision is made.

Thus, the *maximal error probability* of such a code is

$$P_e(W, V) = \max_{1 \leq j \leq N} \left(1 - p(V^{(j)} | w^{(j)}) \right), \quad (4.30)$$

where $p(V^{(j)} | w^{(j)}) = P(Y^n \in V^{(j)} | X^n = w^{(j)})$ – are the probabilities of correct decisions. The *mean error probability* is equal to

$$\bar{P}_e(W, V) = \frac{1}{N} \sum_{j=1}^N \left(1 - p(V^{(j)} | w^{(j)}) \right) \leq P_e(W, V). \quad (4.31)$$

The following useful statement implies that criteria for information transmission based on the average error probability $\bar{P}_e(W, V)$ and on the maximal error probability $P_e(W, V)$ are asymptotically the same.

Lemma 4.11. *For any code (W, V) of size $2N$ with average error probability $\bar{P}_e(W, V)$ there exists a subcode (\tilde{W}, \tilde{V}) of size N which has a maximal error probability $P_e(\tilde{W}, \tilde{V}) \leq 2\bar{P}_e(W, V)$.*

Proof. Put $\epsilon = \bar{P}_e(W, V)$ and assume that there are at least $N + 1$ codewords with error probability $p(V^{(j)} | w^{(j)}) \geq 2\epsilon$, so that it would not be possible to construct the required N -code. In this case, the average error probability of the $2N$ -code would be bounded from below, since

$$\bar{P}_e(W, V) > \frac{1}{2N} 2\epsilon(N + 1) > \epsilon,$$

which contradicts the assumption. \square

Denote by $p_e(n, N)$ (resp. by $\bar{p}_e(n, N)$) the maximal (resp. average) error probability minimized over all codes of length n and size N . Then

$$\frac{1}{2} p_e(n, N) \leq \bar{p}_e(n, 2N) \leq p_e(n, 2N). \quad (4.32)$$

Definition 4.12. The number $R \geq 0$ is called the *achievable transmission rate* if

$$\lim_{n \rightarrow \infty} p_e(n, 2^{nR}) = 0.$$

The supremum of all achievable rates is called the *information capacity* C of the channel $p(y|x)$.

Theorem 4.13 (Shannon's Channel Coding Theorem). *For a memoryless channel*

$$C = C_{\text{Shan}}.$$

Thus, the operationally (and asymptotically) defined transmission capacity is equal to the “one-letter” characteristic of the channel as defined by (4.22). The inequality $C \geq C_{\text{Shan}}$ will follow from the statement

$$\lim_{n \rightarrow \infty} p_e(n, 2^{nR}) = 0, \quad \text{if } R < C_{\text{Shan}},$$

which is called the *direct statement* of the coding theorem, while the *weak converse* amounts to

$$\liminf_{n \rightarrow \infty} p_e(n, 2^{nR}) > 0, \quad \text{if } R > C_{\text{Shan}}, \quad (4.33)$$

which implies $C \leq C_{\text{Shan}}$. In fact, one can prove a stronger result

$$\lim_{n \rightarrow \infty} p_e(n, 2^{nR}) = 1, \quad \text{if } R > C_{\text{Shan}}.$$

Proof. The Weak Converse. Due to inequality (4.32), it is sufficient to prove the analog of the statement (4.33) for the average error probability $\bar{p}_e(n, N)$.

Lemma 4.14. (Fano's Inequality) *Let X, Y be two random variables and $\hat{X} = \hat{X}(Y)$ be an estimator of X with error probability $p_e = P(\hat{X}(Y) \neq X)$. In this case,*

$$H(X|Y) \leq h_2(p_e) + p_e \log(|\mathcal{X}| - 1) \leq 1 + p_e \log |\mathcal{X}|. \quad (4.34)$$

Proof. Let E denote an indicator of the estimation error,

$$E = \begin{cases} 0, & \text{if } \hat{X}(Y) = X; \\ 1, & \text{otherwise.} \end{cases} \quad (4.35)$$

Since E is a function of (X, \hat{X}) and thus is fixed for a given value of X, Y , we have $H(X|Y) = H(E, X|Y)$. But (4.13) implies

$$H(EX|Y) = H(E|Y) + H(X|EY). \quad (4.36)$$

Here $H(E|Y) \leq H(E) = h_2(p_e)$ and

$$H(X|EY) = (1 - p_e)H(X|E = 0, Y) + p_e H(X|E = 1, Y) \leq p_e \log(|\mathcal{X}| - 1),$$

where we have used that $H(X|E = 0, Y) = 0$, since $E = 0$ means that, given Y , we know X exactly. The condition $E = 1$ is the same as $X \neq \hat{X}(Y)$, which leaves the possibility for $|\mathcal{X}| - 1$ values of X . Hence, the maximal possible value of the entropy is $\log(|\mathcal{X}| - 1)$. \square

Consider an arbitrary code (W, V) of size N , with codewords $w^{(1)}, \dots, w^{(N)}$ of length n and a decomposition of \mathcal{Y}^n into $N + 1$ decision domains

$$V^{(0)}, V^{(1)}, \dots, V^{(N)} \subset \mathcal{Y}^n.$$

Denote by X^n a random variable taking the values $w^{(1)}, \dots, w^{(N)}$ with equal probabilities $\frac{1}{N}$ and let $\hat{X}^n = \hat{X}(Y^n)$ be an estimate for X^n such that $\hat{X}^n = w^{(j)}$ if $Y^n \in V^{(j)}$. In this case,

$$\begin{aligned} C_n &\geq I(X^n; Y^n) = H(X^n) - H(X^n|Y^n) \\ &\geq \log N (1 - P\{\hat{X}^n \neq X^n\}) - 1, \end{aligned} \quad (4.37)$$

by Fano's inequality, where $P\{\hat{X}^n \neq X^n\} = \bar{P}_e(W, V)$. Substituting $N = 2^{nR}$ and dividing by nR leads to

$$\bar{P}_e(W, V) \geq 1 - \frac{C_n}{nR} - \frac{1}{nR}. \quad (4.38)$$

Taking the minimum of all possible codes, and using the additivity (4.28), we have in the limit $n \rightarrow \infty$

$$\liminf_{n \rightarrow \infty} \bar{p}_e(n, 2^{nR}) \geq 1 - \frac{C_{\text{Shan}}}{R} > 0$$

for $R > C_{\text{Shan}}$.

Proof of the Direct Statement. The main idea, due to Shannon, is to use *random coding*. Take N independent codewords $w^{(1)}, \dots, w^{(N)}$ with the same probability distribution

$$P\{w^{(i)} = (x_1, \dots, x_n)\} = p_{x_1} \cdots p_{x_n}, \quad (4.39)$$

where the input distribution $P = \{p_x\}$ is chosen such that it maximizes $I(X; Y)$. Note that we have $2^{nH(X)}$ ($2^{nH(Y)}$) typical codewords for the input (output) and *on average* $2^{nH(Y|X)}$ typical outputs for every input codeword w . In order for the error in the discrimination between different words at the output to tend to zero, it is necessary that the sets of typical output words corresponding to different input words do not intersect asymptotically. Hence, the size of the code is

$$N \approx \frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{n(H(Y) - H(Y|X))} = 2^{nI(X; Y)} = 2^{nC_{\text{Shan}}}. \quad (4.40)$$

To make this argument precise, let us call an output word y^n *conditionally typical* for the input word $w = x^n$, if

$$2^{-n(H(Y|X) + \delta)} < p(y^n|w) < 2^{-n(H(Y|X) - \delta)}. \quad (4.41)$$

Denote by $T_w^{n, \delta} \subset \mathcal{Y}^n$ the subset of all conditionally typical words for the given input word w .

Exercise 4.15. Assuming that w is random, with distribution (4.39), show, by using the Law of Large Numbers, that $P\{Y^n \in \overline{T_w^{n, \delta}}\} \leq \varepsilon$ for arbitrary $\varepsilon > 0$ and sufficiently large n .

For a given encoding $W = \{w^{(1)}, \dots, w^{(N)}\}$, we construct a special suboptimal decoding. The subsets $T_{w^{(j)}}^{n,\delta}$ may be intersecting for different j . Hence, to obtain nonintersecting decision domains we put

$$V^{(j)} = T_{w^{(j)}}^{n,\delta} \cap \overline{\left(\bigcup_{k:k \neq j} T_{w^{(k)}}^{n,\delta}\right)}. \quad (4.42)$$

Now, if e.g. the word $w^{(1)}$ was transmitted, the error occurs if and only if

$$y^n \in \overline{V^{(1)}} = \overline{T_{w^{(1)}}^{n,\delta}} \cup \left(\bigcup_{k=2}^N T_{w^{(k)}}^{n,\delta}\right). \quad (4.43)$$

We now assume that the codewords $w^{(1)}, \dots, w^{(N)}$ are selected randomly, as described above, i.e. independently, each with distribution (4.39), and let us evaluate the expectation of the mean error probability $\bar{P}_e(W, V)$. This will give us the required estimate, since obviously

$$\bar{p}_e(n, N) \leq E\bar{P}_e(W, V).$$

Denote by $P_w\{B\}$ the probability $P\{Y^n \in B | X^n = w\}$, where B is some, possibly random, subset of \mathcal{Y}^n . Due to the complete symmetry between the codewords,

$$\begin{aligned} E\bar{P}_e(W, V) &= EP_{w^{(1)}}\{\overline{V^{(1)}}\} \\ &= EP_{w^{(1)}}\{\overline{V^{(1)}} \cap T^{n,\delta}(Y)\} + EP_{w^{(1)}}\{\overline{V^{(1)}} \cap \overline{T^{n,\delta}(Y)}\}, \end{aligned} \quad (4.44)$$

where $T^{n,\delta}(Y)$ is the set of the output δ -typical sequences y^n , defined similarly to Definition 4.2 with the replacement of $H(X)$ by $H(Y)$. Then, taking into account (4.43),

$$E\bar{P}_e(W, V) \leq P\{Y^n \in \overline{T_{w^{(1)}}^{n,\delta}}\} + \sum_{k=2}^N EP_{w^{(1)}}\{T_{w^{(k)}}^{n,\delta} \cap T^{n,\delta}(Y)\} + P\{\overline{T^{n,\delta}(Y)}\}. \quad (4.45)$$

The first and third terms are $\leq \varepsilon$ for sufficiently large n by the corresponding properties of the (conditionally) non-typical words. Each term in the second sum is evaluated as

$$\begin{aligned} &\sum_{w^{(1)}} \sum_{w^{(k)}} \sum_{y^n \in T_{w^{(k)}}^{n,\delta} \cap T^{n,\delta}(Y)} p(y^n | w^{(1)}) p_{w^{(1)}} p_{w^{(k)}} \\ &\leq 2^{n(H(Y|X)+\delta)} \sum_{w^{(1)}} \sum_{w^{(k)}} \sum_{y^n \in T^{n,\delta}(Y)} p(y^n | w^{(1)}) p_{w^{(1)}} p(y^n | w^{(k)}) p_{w^{(k)}} \\ &= 2^{n(H(Y|X)+\delta)} \sum_{y^n \in T^{n,\delta}(Y)} (p_{y^n})^2 \leq 2^{n(H(Y|X)-H(Y)+2\delta)}. \end{aligned}$$

Here, the first inequality is obtained by introducing the factor

$$2^{n(H(Y|X)+\delta)} p(y^n|w^{(k)}) > 1$$

on the set $T_{w^{(k)}}^{n,\delta}$, while the last one follows from the second inequality in the definition of a typical word y^n and the fact that $P\{T^{n,\delta}(Y)\} \leq 1$. Finally,

$$E\bar{P}_e(W, V) \leq 2\varepsilon + (N-1)2^{n(H(Y|X)-H(Y)+2\delta)} \leq 2\varepsilon + N2^{-n(C_{\text{Shan}}-2\delta)},$$

which can be made less than 3ε if $N = 2^{nR}$, with $R < C_{\text{Shan}}$ and δ sufficiently small. \square

It should be noted that the method of random encoding that allows us to prove the existence of asymptotically optimal codes and which discloses their nature is unfortunately not very suitable for practical applications. Its realization even for moderate values of n requires a huge (exponentially growing) amount of computation for the decoding. The development of practically acceptable methods of encoding and decoding is the subject of a special topic in information theory – coding theory – which makes ample use of the methods of modern algebra and combinatorics.

4.5 Wiretap channel

Another important issue in modern information theory is the study of multiple user systems – networks – such as Internet. A specific example of a system with a malicious party is the wiretap channel, which we will briefly discuss here, since it will be used in Ch. 10, which is devoted to quantum information transmission.

Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite alphabets, where alphabet \mathcal{X} is associated with the transmitter, \mathcal{Y} with the receiver, and \mathcal{Z} with an eavesdropper. A *wiretap channel* is determined by a pair of channels $p(y|x); q(z|x)$, one from transmitter to receiver and another from transmitter to eavesdropper. The transmitter sends the words $w = (x_1, \dots, x_n)$ of length n through the channels $p(y^n|x^n), q(z^n|x^n)$, and the goal is to achieve asymptotically exact transmission of a maximal amount of information to the receiver, subject to the condition that the eavesdropper receives an asymptotically negligible amount of information.

Definition 4.16. For a composite wiretap channel, the code (\mathcal{M}, r, V) of size N consists of a collection of messages $\mathcal{M} = \{1, \dots, N\}$, the transition probability $r(x^n|m)$ from \mathcal{M} to \mathcal{X}^n , defining the randomized encoding of the messages $m \in \mathcal{M}$ into words x^n , and a decoding V for the receiver, given by a decomposition of the set \mathcal{Y}^n into $N+1$ pairwise nonintersecting subsets $V^{(0)}, V^{(1)}, \dots, V^{(N)} \subset \mathcal{Y}^n$.

The maximal error probability of such a code is equal to

$$P_e(\mathcal{M}, r, V) = \max_{1 \leq m \leq N} \left(1 - p(V^{(m)}|m) \right), \quad (4.46)$$

where $p(V^{(m)}|m) = \sum_{x^n} P(Y^n \in V^{(m)}|x^n)r(x^n|m)$ is the probability of a correct decision for the receiver. We call R *achievable rate* for the wiretap channel if there exists a sequence of codes $(\mathcal{M}^{(n)}, r^{(n)}, V^{(n)})$ of sizes $N = 2^{nR}$ such that

$$\lim_{n \rightarrow \infty} P_e(\mathcal{M}^{(n)}, r^{(n)}, V^{(n)}) = 0$$

and

$$\lim_{n \rightarrow \infty} I(M^n; Z^n) = 0,$$

where M^n is a random variable, uniformly distributed over the set of messages $\mathcal{M}^{(n)}$. The last condition expresses the requirement that the eavesdropper's information about the transmitted messages should tend to zero. The supremum of the achievable rates is called the *private capacity* C_p of the wiretap channel.

The *Coding Theorem* for the wiretap channel provides the following expression

$$C_p = \max [I(M; Y) - I(M; Z)], \quad (4.47)$$

where the maximum is taken over all possible triples of random variables M, Y, Z , such that the sequence $M, X, (Y, Z)$ is a Markov chain, while the couples X, Y and X, Z are related correspondingly via channels $p(y|x)$ and $q(z|x)$.

The proof is based on the following idea. Let us fix $\delta, \varepsilon > 0$ and a distribution for X and let $T^{n,\delta}$ be the set of δ -typical words of length n . Consider the "pruned" random encoding $W^{(n)}$ which uses N independent words, each *uniformly distributed* over the set $T^{n,\delta}$. Let $V^{(n)}$ be the suboptimal decoding for Y constructed as in (4.42). A modification of the proof of Theorem 4.13 allows us to prove that for $N = 2^{n[I(X;Y)-\delta]}$ the inequality

$$\bar{P}_e(W^{(n)}, V^{(n)}) \leq \varepsilon \quad (4.48)$$

holds with a high probability.

To make the code secret, the transmitter has to sacrifice $n[I(X;Z) + \delta/2]$ bits of information by additional randomization of the messages. Assuming $I(X;Y) - I(X;Z) > 0$, let

$$N_Z = 2^{n[I(X;Z)+\delta/2]}, \quad N_Y = 2^{n[I(X;Y)-I(X;Z)-3\delta/2]},$$

so that $N_Z N_Y = N$. Let us arrange the array of all codewords $W^{(n)}$ as a matrix with N_Y rows and N_Z columns. In this case,

$$W^{(n)} = \{w^{mj}; m = 1, \dots, N_Y; j = 1, \dots, N_Z\}.$$

Next, let for any message m the transmitter choose the values j randomly, with equal probabilities. It can be shown that such a randomization makes almost all transmitted information hidden for the eavesdropper. For every m the collection of codewords

$$\{w^{mj}; j = 1, \dots, N_Z\} \quad (4.49)$$

with high probability contains almost maximal possible information $I(X; Z)$, subject to the condition that the eavesdropper uses the optimal decoding. In other words, for every value of m , the sets of conditionally typical words corresponding to (4.49) cover almost the whole set of typical words for Z . The mutual information between the collections of codewords with different values of m is close to zero, so that randomization inside each collection erases almost all information going to the eavesdropper. This argument, together with relation (4.48), shows that the rate $I(X; Y) - I(X; Z) - \delta$ should be achievable, while randomization is equivalent to using an additional channel $M \rightarrow X$. A modification of a similar argument for any triple M, X, YZ satisfying the conditions of the Coding Theorem allows us to show that $I(M; Y) - I(M; Z) - \delta$ is also an achievable rate.

4.6 Gaussian channel

Consider the continuous analog of the memoryless channel (4.27), with the real line \mathbb{R} as the input and output alphabets. The channel is defined by the transition probability density

$$p(y_i | x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y_i - x_i)^2}{2\sigma^2} \right]; \quad i = 1, \dots, n.$$

Equivalently, one obtains the output random variables Y_i by addition of the noise given by independent identically distributed Gaussian random variables Z_i to the input signal x_i , with $EZ_i = 0$ and $EZ_i^2 = \sigma^2$

$$Y_i = x_i + Z_i; \quad i = 1, \dots, n.$$

Clearly, the naturally defined information capacity of such a channel is infinite if one does not introduce constraints on the signal. Usually, the quadratic constraint is considered

$$\sum_{i=1}^n x_i^2 \leq ns^2, \quad (4.50)$$

motivated by finiteness of the signal power. Under this constraint the channel capacity is computed as

$$C = \frac{1}{2} \log \left(1 + \frac{s^2}{\sigma^2} \right). \quad (4.51)$$

This formula also belongs to Shannon and it also has a simple heuristic explanation.

The inequalities

$$\frac{1}{n} \sum_{i=1}^n Z_i^2 \leq \sigma^2 + \varepsilon; \quad \frac{1}{n} \sum_{i=1}^n Y_i^2 \leq s^2 + \sigma^2 + \varepsilon$$

hold with high probability for arbitrary $\varepsilon > 0$ and sufficiently large n .

Exercise 4.17. Prove this statement by using the Markov inequality $P\{X \geq \varepsilon\} \leq EX/\varepsilon$ for a nonnegative random variable X , and the relation (4.50).

The first of these inequalities means that, for any input word $w = (x_1, \dots, x_n)$, the output vector $Y^n = (Y_1, \dots, Y_n)$ lies in an n -dimensional ball with radius $\sqrt{n(\sigma^2 + \varepsilon)}$ and center w , while the second tells us that the output vectors for arbitrary inputs that satisfy constraint (4.50) lie in an n -dimensional ball of radius $\sqrt{n(s^2 + \sigma^2 + \varepsilon)}$, with its center at the origin. The ratio of the volumes of these balls gives an approximate value for the quantity of balls of the first type within the big ball,

$$N = \left[\frac{\sqrt{n(s^2 + \sigma^2 + \varepsilon)}}{\sqrt{n(\sigma^2 + \varepsilon)}} \right]^n = 2^{nR},$$

where $R = \frac{1}{2} \log \left(1 + \frac{s^2}{\sigma^2 + \varepsilon} \right)$, which can be made arbitrarily close to (4.51).

4.7 Notes and references

1. The ideas of reliable data transmission have been promoted since the 1930s and even earlier (see Verdú [211] for a detailed historical survey). Remarkably, Kotelnikov, the Russian pioneer in the field, throughout his life maintained a deep interest in the foundations of quantum mechanics. The mathematical tradition in information theory goes back to the 1950s, when Khinchin and Kolmogorov gave insights into Shannon's discoveries from the viewpoint of advanced probability theory. Nowadays, there are many excellent texts on information theory. Here, we mainly follow the book of Cover and Thomas [42], which provides an enlightening, nontechnical introduction to the subject. The method of types was systematically developed by Csiszar and Körner [44].
2. The idea of using the exponential function in the Markov inequality with subsequent large deviation estimates goes back to Bernstein, followed by Cramér. Its importance to information theory and statistics was emphasized in the work of Chernoff, see e.g. [42], lemma 12.9.1, and also [66].
3. For a complete proof of Exercise 4.8 and its application to the additivity of the Shannon capacity, see Gallager [66], theorems 4.4.1, 4.5.1.
4. The heuristic argument for the direct coding theorem was already present in Shannon's pioneering paper [182]. Our proof differs from the one given in the book [42] in that it uses conditional rather than joint typicality. It is this approach that allows for the noncommutative extension, see Section 5.6.

5. Theory of multiuser systems is presented in the books of Csiszar and Körner [44], and Cover and Thomas [42]. The proof of the coding theorem for the wiretap channel is given in [44].
6. For a rigorous proof of the formula (4.51), see e.g. [42]. The quantum analog of the channel with additive Gaussian noise is considered in Section 12.1.4.

Chapter 5

The classical-quantum channel

5.1 Codes and achievable rates

As was explained in Section 3.3.1, a simple mathematical model of a quantum communication channel is given by a map $x \rightarrow S_x$ that transforms the letters x of a (finite) input alphabet \mathcal{X} into the quantum states at the channel output. We will call such a model a *classical-quantum* (c-q) channel. If an observable $M = \{M_y\}$ is measured at the output of such a channel, the conditional probability to obtain the outcome y , under the condition that the signal x was sent, is given by the formula

$$p(y|x) = \text{Tr } S_x M_y. \quad (5.1)$$

Thus, for a fixed quantum measurement we have a usual classical channel. This leads to the question as to the maximum amount of classical information that can be transmitted through the quantum communication channel with asymptotically vanishing error, i.e. its “classical capacity”. This question will be considered in detail in the present chapter.

Consider the composite c-q channel which maps a word $w = (x_1, \dots, x_n)$ into the product state $S_w = S_{x_1} \otimes \dots \otimes S_{x_n}$ in the space $\mathcal{H}^{\otimes n}$. The process of transmission of classical information is described by the diagram

$$i \xrightarrow{\text{encoding}} w \xrightarrow{\text{channel}} S_w \xrightarrow{\text{decoding}} j \quad (5.2)$$

The assumption that a word w is mapped into the tensor product of states S_{x_j} corresponds to the definition of a memoryless channel in the classical case. At the channel output, there is a receiver who performs a measurement of an observable $M = \{M_j^{(n)}\}$ in the space $\mathcal{H}^{\otimes n}$ (the outcome j means that the receiver decides that the signal j was sent). Thus, the resolution of the identity in the space $\mathcal{H}^{\otimes n}$ describes the overall statistics of the decision procedure, including both the physical measurement and the subsequent classical information processing. The choice of the observable M is formally similar to the choice of the decision rule in the classical case, but here plays a much more significant role, as we shall see.

Definition 5.1. A *code* (W, M) of size N consists of a classical *codebook* $W = \{w^{(i)}; i = 1, \dots, N\}$, which is a collection of N codewords of length n , and a quantum *decision rule* given by an observable $M = \{M_j; j = 0, 1, \dots, N\}$ in $\mathcal{H}^{\otimes n}$. An outcome $j = 1, \dots, N$ corresponds to the decision that the word w^j was sent, while $j = 0$ means no definite decision.

The probability of decoding the signal i as j is equal to

$$p_{WM}(j|i) = \text{Tr } S_{w(i)} M_j, \quad j = 0, 1, \dots, N. \quad (5.3)$$

In this case, the probability of a correct decision is $p_{WM}(i|i) = \text{Tr } S_{w(i)} M_i$, while the maximal error probability of the code (W, M) is equal to

$$P_e(W, M) = \max_{1 \leq j \leq N} [1 - p_{WM}(j|j)]. \quad (5.4)$$

The average error probability is

$$\bar{P}_e(W, M) = \frac{1}{N} \sum_{j=1}^N [1 - p_{WM}(j|j)]. \quad (5.5)$$

In what follows we denote by

$$p_e(n, N) = \min_{W, M} P_e(W, M), \quad \bar{p}_e(n, N) = \min_{W, M} \bar{P}_e(W, M), \quad (5.6)$$

the maximal and the average error, resp., minimized over all codes (W, M) of size N , using words of length n .

Definition 5.2. The *classical capacity* C of the c-q channel $x \rightarrow S_x$ is equal to the supremum of the transmission rates R that are achievable, in the sense that $\lim_{n \rightarrow \infty} p_e(n, 2^{nR}) = 0$ or, equivalently (by inequalities (4.32)), that $\lim_{n \rightarrow \infty} \bar{p}_e(n, 2^{nR}) = 0$.

5.2 Formulation of the coding theorem

The notion of the entropy of a quantum state is a natural extension of the entropy of a probability distribution. Let S be a density operator in d -dimensional Hilbert space \mathcal{H} , and let

$$S = \sum_{j=1}^d s_j |e_j\rangle\langle e_j|$$

be its spectral decomposition. Its eigenvalues s_j form a probability distribution. The *von Neumann entropy* of the density operator S is defined as

$$H(S) = - \sum_{j=1}^d s_j \log s_j = \text{Tr } \eta(S), \quad (5.7)$$

where the function $\eta(\cdot)$ is given by (4.2), and is equal to the Shannon entropy of the probability distribution $\{s_j\}$. In what follows we use a more spectacular expression

$H(S) = -\text{Tr } S \log S$, although the logarithm is not defined for degenerate density operators. As in the case of distributions, this will not lead to ambiguities. From (4.3) it follows that

$$0 \leq H(S) \leq \log d,$$

with the minimum achieved on pure states, and the maximum on the chaotic state $\bar{S} = \frac{1}{d}I$. As in the classical case, the entropy is a measure of the uncertainty and also of the information content of the state (this last statement will be substantiated in Section 5.5).

The following properties of the quantum entropy are easy to check:

Exercise 5.3.

- i. *Unitary invariance:* $H(VSV^*) = H(S)$, where V is a unitary operator. Moreover, this equality holds for any operator V , isometric on the support of S ;
- ii. *Additivity:* $H(S_1 \otimes S_2) = H(S_1) + H(S_2)$.

In earlier works on quantum communications, the quantity

$$\chi(\{\pi_x\}; \{S_x\}) = H\left(\sum_x \pi_x S_x\right) - \sum_x \pi_x H(S_x) \quad (5.8)$$

with $\pi = \{\pi_x\}$ a probability distribution on \mathcal{X} , was used to evaluate the capacity of the c-q channel $x \rightarrow S_x$ on heuristic grounds. The quantity (5.8) can be considered a formal quantum analog of the Shannon information, where the first term plays the role of the output entropy, while the second – that of the conditional entropy of the output with respect to the input. Remarkably, this quantity appears to be strictly related to the classical capacity of the c-q channel as defined in the previous section. When the states $\{S_x\}$ are fixed and there is no risk of confusion, we shall abbreviate the notation (5.8) as $\chi(\pi)$ and put

$$C_\chi = \max_\pi \chi(\pi). \quad (5.9)$$

The main goal of the next sections will be the proof of the following Coding Theorem:

Theorem 5.4 (Holevo [97]; Schumacher and Westmoreland [177]). *The classical capacity C of the c-q channel $x \rightarrow S_x$, defined as in Definition 5.2, is equal to C_χ .*

Note that the quantum entropy is continuous on the compact set of quantum states, hence χ is continuous in π and the maximum in the above formula is indeed achieved. Let us introduce the notation

$$\bar{S}_\pi = \sum_x \pi_x S_x \quad (5.10)$$

for the average output state. In general,

$$\chi(\pi) \leq H(\bar{S}_\pi),$$

with equality in the case of the pure-state channel $S_x = |\psi_x\rangle\langle\psi_x|$. Since the maximum possible value of the entropy is $\log \dim \mathcal{H}$, this theorem implies an absolute upper bound on the classical capacity:

$$C \leq \log \dim \mathcal{H}. \quad (5.11)$$

Thus, in spite of the fact that in a Hilbert space there are infinitely many different pure states, the finite-dimensional quantum carrier cannot be used to transmit an unlimited amount of classical information. The upper bound (5.11) is achieved for the channel with orthogonal pure states $S_x = |e_x\rangle\langle e_x|$; $x = 1, \dots, d$, where $\{e_x\}$ is an orthonormal basis in \mathcal{H} , $d = \dim \mathcal{H}$, and for the uniform distribution $\pi_x \equiv 1/d$. Note that such states, as a rule, cannot be obtained at the output of a realistic communication channel. It is remarkable, however, that orthogonality of the output states is not necessary in order to attain the capacity of the ideal quantum channel.

Example 5.5. Consider the *trine channel* with the three equiangular pure states ψ_0, ψ_1, ψ_2 in a two-dimensional Hilbert space as in Example 2.27. Taking the equal probabilities $\pi_x = \frac{1}{3}$, we obtain the chaotic average output state

$$\bar{S}_\pi = \sum_{x=0}^2 \pi_x |\psi_x\rangle\langle\psi_x| = \frac{1}{2} I, \quad (5.12)$$

implying $C_\chi = \log 2 = 1$ bit, i.e. the capacity of the ideal channel.

More generally, if $\{\psi_x\}$ is an arbitrary overcomplete system in a Hilbert space \mathcal{H} , the pure state channel $x \rightarrow \frac{|\psi_x\rangle\langle\psi_x|}{\langle\psi_x|\psi_x\rangle}$ has the maximum possible capacity $\log d$, and the maximum is attained on the distribution $\pi_x = \langle\psi_x|\psi_x\rangle/d$.

Example 5.6. Consider a binary channel with two pure states with unit vectors ψ_0, ψ_1 as in Example 2.25. In this case,

$$C_\chi = h_2\left(\frac{1+\epsilon}{2}\right), \quad (5.13)$$

where $\epsilon = |\langle\psi_0|\psi_1\rangle|$. Indeed, the entropy $H(\bar{S}_\pi)$ is maximized by the uniform distribution $\pi_0 = \pi_1 = \frac{1}{2}$ since, by concavity of the quantum entropy (see Corollary 7.17 in Chapter 7),

$$H\left(\frac{1}{2}|\psi_0\rangle\langle\psi_0| + \frac{1}{2}|\psi_1\rangle\langle\psi_1|\right) = H\left(\frac{S+S'}{2}\right) \geq \frac{1}{2}(H(S) + H(S')),$$

where

$$S = \pi_0 |\psi_0\rangle\langle\psi_0| + \pi_1 |\psi_1\rangle\langle\psi_1|, \quad S' = \pi_1 |\psi_0\rangle\langle\psi_0| + \pi_0 |\psi_1\rangle\langle\psi_1|,$$

and $H(S') = H(S)$ due to the reflection symmetry with respect to the bisector of the angle between the two vectors. Similar to the solution of Exercise 2.26, the eigenvalues of the density operator $\frac{1}{2} |\psi_0\rangle\langle\psi_0| + \frac{1}{2} |\psi_1\rangle\langle\psi_1|$ are equal to $\frac{1 \pm \epsilon}{2}$, whence (5.13) follows.

5.3 The upper bound

Consider an arbitrary c-q channel $x \rightarrow S_x$ and an observable $M = \{M_y\}$ at its output. Note that the alphabets \mathcal{X} and \mathcal{Y} need not coincide. The variables x and y are related by the classical channel $p_M(y|x) = \text{Tr } S_x M_y$. Denote by

$$\mathcal{I}_1(\pi, M) = \sum_{xy} \pi_x p_M(y|x) \log \left(\frac{p_M(y|x)}{\sum_{x'} p_M(y|x') \pi_{x'}} \right) \quad (5.14)$$

the Shannon information between the variables x and y corresponding to some input distribution $\pi = \{\pi_x\}$.

Exercise 5.7. For given π and \mathcal{Y} the information quantity $\mathcal{I}_1(\pi, M)$ is a continuous and convex function of the transition probability $p_M(y|x)$, and hence of M . Hint: use the corresponding property of the Shannon information $I(X; Y)$.

For given distribution π on the input alphabet, consider the *accessible information*

$$A(\pi) = \sup_M \mathcal{I}_1(\pi, M),$$

where the supremum is taken over all possible observables at the output of the channel $x \rightarrow S_x$.

Proposition 5.8 (Davies [47]). *The supremum of the information quantity $\mathcal{I}_1(\pi, M)$ is attained by the observable M^0 of the form*

$$M_y^0 = |\phi_y\rangle\langle\phi_y|; \quad y = 1, \dots, m, \quad (5.15)$$

where $m \leq d^2$ in the case of complex \mathcal{H} ($m \leq \frac{d(d+1)}{2}$ in the case of real \mathcal{H}).

Proof. For any $k = 2, 3, \dots$ the set \mathfrak{M}_k of observables with k outcomes is compact, and the continuous convex functional $\mathcal{I}_1(\pi, M)$ attains its maximum on the set \mathfrak{M}_k . Assume that \hat{M}^0 maximizes $\mathcal{I}_1(\pi, M)$ on \mathfrak{M}_k . By throwing off, if necessary, zero components, we obtain the observable $\tilde{M}^0 \in \mathfrak{M}_l, l \leq k$, for which

$\mathcal{I}_1(\pi, \tilde{M}^0) = \mathcal{I}_1(\pi, \hat{M}^0)$. By making the spectral decomposition of the components of the observable \tilde{M}^0 into rank one operators, we can construct the new observable M^0 of the form (5.15), with $m \geq l$, for which \tilde{M}^0 is a coarse-graining,

$$\underbrace{M_1, M_2, \dots}_{=\tilde{M}_1}, \underbrace{\dots}_{=\tilde{M}_2}, \dots, \underbrace{\dots}_{=\tilde{M}_l}, M_m.$$

Then $\mathcal{I}_1(\pi, M^0) \geq \mathcal{I}_1(\pi, \tilde{M}^0)$. In fact, this is a statement about the classical Shannon information, $I(X; f(Y)) \leq I(X; Y)$ or, equivalently, $H(X|f(Y)) \geq H(X|Y)$, which follows from the monotonicity of the classical conditional entropy, since $H(X|Y) = H(X|f(Y), Y)$.

The function $M \rightarrow \mathcal{I}(\pi, M)$ is convex. Therefore, we can assume that the maximizing observable $M^0 = \{M_y^0\}$ is an extreme point of the set \mathfrak{M}_m . Then, according to Theorem 2.21, the operators M_y^0 are linearly independent and Exercise 1.5 implies the estimate $m \leq d^2$. Since k was arbitrary, it follows that $\sup_M \mathcal{I}_1(\pi, M)$ is attained on an observable from the compact convex set \mathfrak{M}_{d^2} . \square

Theorem 5.9 (Holevo [95]). *For an arbitrary distribution π*

$$A(\pi) \equiv \max_M \mathcal{I}_1(\pi, M) \leq \chi(\pi), \quad (5.16)$$

with equality attained if and only if the operators $\pi_x S_x; x \in \mathcal{X}$ all commute.

In Chapter 7, we will obtain the information bound (5.16) as a corollary of the rather general monotonicity property of the quantum relative entropy. Here we give instead an outline of the original proof, based on the comparison of the convexity of the classical and quantum entropies. This proof reveals a connection between the quantum entropy and the noncommutative analog of the Fisher information, and moreover allows us to obtain a necessary and sufficient condition for equality. The latter condition will be used in the next section, in order to establish the nonclassical property of superadditivity of the Shannon information for the c-q channel.

Proof. Without loss of generality, we assume all $\pi_x > 0$. In this case, the condition for equality amounts to the commutativity of all operators S_x . Now, the equality in (5.16) is apparently achieved by the common spectral measure M of the operators S_x .

To prove the inequality let us first consider the case of two states S_0, S_1 . Denote

$$\chi(t) = H((1-t)S_0 + tS_1) - (1-t)H(S_0) - tH(S_1), \quad t \in [0, 1]. \quad (5.17)$$

Also set $S_t = (1-t)S_0 + tS_1$, $D = S_1 - S_0$ and let $S_t = \sum_k s_k E_k$ be the spectral decomposition of the operator S_t . Here and in what follows, s_k will denote *strictly positive* eigenvalues.

By using Cauchy's integral formula, we have

$$S_t \log S_t = \frac{1}{2\pi i} \oint (Iz - S_t)^{-1} z \log z dz,$$

where $z \log z$ is the branch of the function, analytic in the right halfplane, and the integral is taken over a closed contour embracing the segment $[\varepsilon, 1]$, $\varepsilon > 0$, which contains all positive eigenvalues of S_t . Differentiating twice with respect to t ,

$$[S_t \log S_t]'' = \frac{1}{2\pi i} \oint [(Iz - S_t)^{-1}]'' z \log z dz,$$

where, using the resolvent expansion and taking into account that $S_t' = D$,

$$[(Iz - S_t)^{-1}]'' = 2(Iz - S_t)^{-1} D (Iz - S_t)^{-1} D (Iz - S_t)^{-1}.$$

By using the fact that $\text{Tr } E_k D E_j D = \text{Tr } E_j D E_k D$ due to (1.11), we have

$$\begin{aligned} \chi''(t) &= -\frac{1}{\pi i} \oint \text{Tr } (Iz - S_t)^{-2} D (Iz - S_t)^{-1} D z \log z dz \\ &= -\frac{1}{\pi i} \oint \sum_{k,j} \frac{\text{Tr } E_k D E_j D}{(z - s_k)^2 (z - s_j)} z \log z dz \\ &= -\sum_{k,j} (\text{Tr } E_k D E_j D) f(s_k, s_j), \end{aligned} \quad (5.18)$$

where

$$\begin{aligned} f(a, b) &= \frac{1}{2\pi i} \oint \left[\frac{1}{(z - a)^2 (z - b)} + \frac{1}{(z - b)^2 (z - a)} \right] z \log z dz \\ &= \frac{\log a - \log b}{a - b}, \quad a \neq b; \quad f(a, a) = a^{-1}. \end{aligned} \quad (5.19)$$

Note that $\text{Tr } E_k D E_j D = \text{Tr } E_k D E_j D E_k \geq 0$. By using the inequality

$$f(a, b) \geq \frac{2}{a + b}, \quad 0 < a \leq 1, \quad 0 < b \leq 1, \quad (5.20)$$

in which equality is obtained if and only if $a = b$, we have

$$\chi''(t) \leq -\sum_{k,j} \text{Tr } E_k D E_j D \frac{2}{s_k + s_j}, \quad (5.21)$$

with equality attained if and only if $\text{Tr } E_k D E_j D = 0$ for all $k \neq j$. The latter condition is equivalent to $[D, S_t] = 0$, i.e. $[S_0, S_1] = 0$, because of the identity

$$\text{Tr } [D, S_t]^* [D, S_t] = \sum_{k,j} (s_k - s_j)^2 \text{Tr } E_k D E_j D.$$

Exercise 5.10. Prove relation (5.19) and inequality (5.20).

Exercise 5.11. Show that the operator

$$L_t = \sum_{k,j} E_k D E_j \frac{2}{s_k + s_j}$$

is a solution of the equation

$$S_t \circ L_t \equiv \frac{1}{2}[S_t L_t + L_t S_t] = D.$$

Moreover,

$$\sum_{k,j} \text{Tr } E_k D E_j D \frac{2}{s_k + s_j} = \text{Tr } D L_t = \text{Tr } S_t L_t^2. \quad (5.22)$$

The operator L_t is a noncommutative (symmetric) analog of the logarithmic derivative of the operator-valued function S_t , while (5.22) is an analog of the Fisher information quantity in mathematical statistics. From (5.21), (5.22) it follows that

$$\chi''(t) \leq -\text{Tr } D L_t = -\text{Tr } S_t L_t^2, \quad 0 < t < 1, \quad (5.23)$$

in particular $\chi''(t) \leq 0$, so that $\chi(t)$ is concave on the segment $[0, 1]$ with $\chi(0) = \chi(1) = 0$. Moreover, equality in (5.23) is achieved if and only if $[S_0, S_1] = 0$.

Now let $M = \{M_y\}$ be an arbitrary observable, let $P_t(y) = \text{Tr } S_t M_y = (1-t)P_0(y) + tP_1(y)$ be its distribution in the state S_t , and $J_M(t) = \mathcal{J}_1(\pi, M)$, where $\pi = \{1-t, t\}$. In other words,

$$J_M(t) = H((1-t)P_0 + tP_1) - (1-t)H(P_0) - tH(P_1),$$

where $H(P)$ is the Shannon entropy of the probability distribution P . Also denote $D(y) = P_1(y) - P_0(y)$. Applying the previous argument to the diagonal matrix $\text{diag}[P_t(y)]$ instead of the state S_t , we obtain

$$J_M''(t) = -\sum_y \frac{D(y)^2}{P_t(y)} = -\text{Tr } D \Lambda_t, \quad (5.24)$$

where

$$\Lambda_t = \sum_y M_y \frac{D(y)}{P_t(y)}.$$

The first equality in (5.24) holds due to commutativity of the diagonal matrices. The right-hand side of (5.24) is equal to minus the classical Fisher information for the

family of measurement probability distributions. We have $J_M''(t) \leq 0$, so that the function $J_M(t)$ is concave on $[0, 1]$, and $J_M(0) = J_M(1) = 0$.

Finally, let us prove the inequality between the classical and the quantum Fisher informations

$$\text{Tr } D\Lambda_t \leq \text{Tr } DL_t, \quad (5.25)$$

which implies, via (5.23), (5.24), that $J_M''(t) \geq \chi''(t)$, and hence

$$J_M(t) \leq \chi(t), \quad 0 < t < 1, \quad (5.26)$$

with the sharp inequality if $[S_0, S_1] \neq 0$.

Proof of the inequality (5.25). First, note that

$$\text{Tr } D\Lambda_t = \sum_y \frac{D(y)^2}{P_t(y)} \geq \text{Tr } S_t \Lambda_t^2, \quad (5.27)$$

since

$$\Lambda_t^2 \leq \sum_y M_y \left[\frac{D(y)}{P_t(y)} \right]^2$$

by the following operator generalization of the Cauchy–Schwarz inequality:

Exercise 5.12. Prove the following statement: given a resolution of the identity $\{M_y\}$, one has for any real c_y

$$\left(\sum_y c_y M_y \right)^2 \leq \sum_y c_y^2 M_y.$$

Now by using (5.27), we obtain

$$\begin{aligned} \text{Tr } DL_t &= \text{Tr } S_t L_t^2 = \text{Tr } S_t [\Lambda_t + (L_t - \Lambda_t)]^2 \geq \text{Tr } S_t \Lambda_t^2 + 2\text{Tr } S_t (L_t - \Lambda_t) \circ \Lambda_t \\ &= -\text{Tr } S_t \Lambda_t^2 + 2\text{Tr } (S_t \circ L_t) \Lambda_t = 2\text{Tr } D\Lambda_t - \text{Tr } S_t \Lambda_t^2 \\ &= \text{Tr } D\Lambda_t + [\text{Tr } D\Lambda_t - \text{Tr } S_t \Lambda_t^2] \geq \text{Tr } D\Lambda_t. \end{aligned}$$

□

This completes the proof of Theorem 5.9 in the case of two states. The case of several states $S_x; x = 0, 1, \dots, k$, with distribution $\pi = \{\pi_x; x = 0, 1, \dots, k\}$ is reduced to the case of two states with the help of the following recurrent formula:

Exercise 5.13.

$$\chi(\pi) = \sum_{m=1}^k (\pi_0 + \dots + \pi_m) \chi_m(t_m),$$

where

$$\begin{aligned}\chi_m(t) &= H((1-t)S_0^m + tS_m) - (1-t)H(S_0^m) - tH(S_m), \\ t_m &= \frac{\pi_m}{\pi_0 + \cdots + \pi_m}, \\ S_0^m &= \sum_{j=0}^{m-1} \frac{\pi_j S_j}{\pi_0 + \cdots + \pi_{m-1}}.\end{aligned}$$

Similarly

$$\mathcal{J}_1(\pi, M) = \sum_{m=1}^k (\pi_0 + \cdots + \pi_m) J_M^m(t_m),$$

where $J_M^m(t_m)$ are defined in the same way as $\chi_m(t_m)$, via the Shannon entropies of the measurement probability distributions. From the proven result for two states, $J_M^m(t_m) \leq \chi_m(t_m)$. Hence, the inequality (5.16) follows. If there is at least a pair of noncommuting states, we can take them as S_0, S_1 , and the sharp inequality follows. \square

5.4 Proof of the weak converse

The bound (5.16) is the main tool for proving the weak converse of the Quantum Coding Theorem, which implies $C \leq C_\chi$.

Theorem 5.14 (Weak Converse). *If $R > C_\chi$, then*

$$\liminf_{n \rightarrow \infty} \bar{p}_e(n, 2^{nR}) > 0. \quad (5.28)$$

Proof. Consider the composite, memoryless c-q channel $w \rightarrow S_w$, where $w = x^{(n)}$ denote words of length n . Let $\mathcal{J}_n(\pi^{(n)}, M^{(n)})$ be the Shannon information defined for this composite channel similarly to $\mathcal{J}_1(\pi, M)$, where $\pi^{(n)} = \{\pi_w\}$ is a probability distribution of codewords of length n and $M^{(n)}$ is an observable in $\mathcal{H}^{\otimes n}$. Then (5.16) implies

$$\mathcal{J}_n(\pi^{(n)}, M^{(n)}) \leq C_\chi^{(n)}, \quad (5.29)$$

where

$$C_\chi^{(n)} = \max_{\pi^{(n)}} \chi(\{\pi_w\}; \{S_w\}) = \max_{\pi^{(n)}} \left[H \left(\sum_w \pi_w S_w \right) - \sum_w \pi_w H(S_w) \right]. \quad (5.30)$$

Lemma 5.15. *The sequence $C_\chi^{(n)}$ is additive, i.e. $C_\chi^{(n)} = nC_\chi$.*

Proof. The inequality $C_\chi^{(n)} \geq nC_\chi$ follows from the additivity of the entropy by using the product input probability distributions $\pi^{(n)}$. The proof of the converse inequality $C_\chi^{(n)} \leq nC_\chi$ follows from subadditivity of the von Neumann entropy with respect to the tensor products (see Corollary 7.4), which implies

$$\chi_n(\pi^{(n)}) \leq \sum_{k=1}^n \chi(\pi^{(n,k)}), \quad (5.31)$$

where $\pi^{(n,k)}$ is the k -th marginal distribution of π on \mathcal{X} . \square

Defining

$$C_n = \max_{\pi^{(n)}, M^{(n)}} \mathcal{I}_n(\pi^{(n)}, M^{(n)}), \quad (5.32)$$

we thus have $C_n \leq nC_\chi$ and, applying the classical Fano's inequality (4.34), we obtain, similar to (4.38),

$$\overline{P}_e(W, M) \geq 1 - \frac{C_n}{nR} - \frac{1}{nR} \geq 1 - \frac{C_\chi}{R} - \frac{1}{nR}, \quad (5.33)$$

and hence, (5.28) follows. \square

The quantity C_n is the maximal Shannon information accessible through n uses of the c-q channel when arbitrary decodings are allowed on the quantum output $\mathcal{H}^{\otimes n}$. The following result clarifies its relation to the classical capacity.

Proposition 5.16. *The classical capacity C of the channel $x \rightarrow S_x$ is equal to*

$$\sup_n \frac{1}{n} C_n = \lim_n \frac{1}{n} C_n. \quad (5.34)$$

Proof. The equality in (5.34) follows from superadditivity of the sequence $\{C_n\}$,

Exercise 5.17. Show that the sequence C_n is superadditive, i.e. $C_{n+m} \geq C_n + C_m$, by taking product probability distributions $\pi^{(n+m)} = \pi^{(n)} \times \pi^{(m)}$.

The inequality $C \leq \sup_n \frac{1}{n} C_n$ follows from the first relation in (5.33). The opposite inequality $C \geq \sup_n \frac{1}{n} C_n$ follows from the direct statement of Shannon's Coding Theorem 4.13 applied to composite channels. Indeed, let us show that any $R < \sup_n \frac{1}{n} C_n$ is an achievable rate. We have $n_0 R < C_{n_0}$ for some n_0 . Hence, we can choose an observable $M^{(n_0)}$ in $\mathcal{H}^{\otimes n_0}$ such that

$$n_0 R < \max_{\pi^{(n_0)}} \mathcal{I}_n(\pi^{(n_0)}, M^{(n_0)}) \equiv C(M^{(n_0)}).$$

The quantity $C(M^{(n_0)})$ is the Shannon capacity of the classical channel

$$p_M^{(n_0)}(j|x^{n_0}) = \text{Tr } S_{x^{n_0}} M_j^{(n_0)}. \quad (5.35)$$

Therefore, the minimal error probability for this channel satisfies $\tilde{p}_e(n, 2^{n(n_0 R)}) \rightarrow 0$ as $n \rightarrow \infty$. Apparently, $\tilde{p}_e(nn_0, 2^{n(n_0 R)}) \leq \tilde{p}_e(n, 2^{n(n_0 R)})$, since the channel (5.35) corresponds to a certain block decoding for the initial c-q channel. Thus, $\tilde{p}_e(nn_0, 2^{n(n_0 R)}) \rightarrow 0$ as $n \rightarrow \infty$. For arbitrary n' , one can find n such that $nn_0 \leq n' \leq (n+1)n_0$. Then

$$\tilde{p}_e(n', 2^{n'R}) \leq \tilde{p}_e(nn_0, 2^{(n+1)n_0 R}) \leq \tilde{p}_e(nn_0, 2^{n(n_0 R')}) \rightarrow 0, \quad (5.36)$$

if R' is chosen such that $R(1 + 1/n) \leq R' < \sup_n \frac{1}{n} C_n$ for sufficiently large n . \square

The quantity

$$C_1 = \max_{\pi, M} J_1(\pi, M)$$

which we call the *Shannon capacity* of the c-q channel $x \rightarrow S_x$ is of special interest. Similar to the proof of Proposition 5.16, it can be shown to be equal to the capacity of the c-q channel $x \rightarrow S_x$, with the additional restriction that only product observables $\{M_{y_1}^1 \otimes \cdots \otimes M_{y_n}^n\}$ are allowed at the output of the composite channel. More generally, we call the decoding $M^{(n)} = \{M_j^{(n)}\}$ in $\mathcal{H}^{\otimes n}$ *unentangled*, if

$$M_j^{(n)} = \sum_{y^n} p(j|y^n) M_{y_1}^1 \otimes \cdots \otimes M_{y_n}^n, \quad (5.37)$$

where $\{M_{y_k}^k\}$ is a quantum observable for the k -th copy of the space \mathcal{H} , and $p(j|y^n)$ is a conditional probability describing the classical (stochastic) transformation of the measurement outcomes of a product observable

$$\{M_{y_1}^1 \otimes \cdots \otimes M_{y_n}^n\} \quad (5.38)$$

at the output.

Exercise 5.18. Denoting by C_{ud} the supremum of achievable rates for the codes (W, M) , with the additional constraint (5.37) on the decoding M , show that $C_{ud} = C_1$. Hint: the effect of the decoding (5.37) can be described by concatenation of the classical composite channel corresponding to the product observable (5.38) and the channel $p(j|y^n)$. Use the data-processing inequality (4.20) and the analog of the first inequality in (5.33) to prove that $C_{ud} \leq C_1$. For the proof of the converse inequality, use the direct statement of the classical coding theorem for the channel $p_M(y|x) = \text{Tr } S_x M_y$ to prove that any rate $R < \max_{\pi} J_1(\pi, M)$ is achievable. Since the decoding M in \mathcal{H} is arbitrary, it follows that any rate $R < C_1$ is achievable. Hence, $C_1 \leq C_{ud}$.

So far, we have shown that

$$C_1 \leq C \leq C_{\mathcal{X}}.$$

The direct statement of the coding theorem, which will be proven below, implies $C = C_\chi$. The following result means that $C_1 < C$ for channels with essentially quantum output. Thus, for the memoryless c-q channels the capacity can be increased by using entangled decodings. It follows that, while $C_n = nC_1$ for a classical memoryless channel (see Proposition 4.9), in the quantum case strict superadditivity of the classical information $C_n > nC_1$ is possible, the reason being the existence of entangled observables at the output of the composite quantum channel. One can say that this is a dual manifestation of the EPR correlations. The latter arise when the state of a composite quantum system is entangled, while the measurements are local, i.e. unentangled. Here, the strict superadditivity of the information appears for the product states and entangled measurements.

Proposition 5.19. *The equality $C_1 = C$ holds if and only if the operators $\pi_x^0 S_x$ commute for a probability distribution π^0 that maximizes $\chi(\pi)$.*

Proof. The sufficiency of the condition is obvious. Conversely, let $C_1 = C$. In this case, by using Proposition 5.8, compactness of the set of probability distributions $\{\pi\}$ and continuity of the Shannon information $\mathcal{I}_1(\pi^1, M^1)$, we have $C_1 = \mathcal{I}_1(\pi^1, M^1)$ for some distribution π^1 and observable M^1 . Hence, $\mathcal{I}_1(\pi^1, M^1) = \chi(\pi^0)$. It follows that $\chi(\pi^1) = \chi(\pi^0)$, otherwise, by using inequality (5.16), we would have

$$\mathcal{I}_1(\pi^1, M^1) \leq \chi(\pi^1) < \chi(\pi^0).$$

Thus, we can replace π^0 with π^1 , $\mathcal{I}_1(\pi^1, M^1) = \chi(\pi^1)$, and necessity follows from the second statement of Theorem 5.9. \square

The following examples (more detailed references and discussion are given in Section 5.8) illustrate the inequality $C_1 < C$.

Example 5.20. For the trine channel,

$$C_1 = 1 - h_2\left(\frac{1 + \sqrt{3}/2}{2}\right) \approx 0.645 \text{ bit} \quad (5.39)$$

which is obtained for the non-uniform distribution $\pi = [1/2, 1/2, 0]$ and for the optimal measurement for the two equiprobable states ψ_0, ψ_1 , see Sasaki et al. [176]. Thus, $C_1 < C$ and the sequence $\{C_n\}$ is strictly superadditive. Note that taking the uniform distribution $\pi = [1/3, 1/3, 1/3]$ and the information optimal observable results in the smaller quantity $\log(3/2) \approx 0.585$ bit.

Example 5.21. The Shannon capacity of the c-q channel with two pure states is given by

$$C_1 = \max_{\pi, M} \mathcal{I}(\pi, M) = 1 - h_2\left(\frac{1 + \sqrt{1 - \epsilon^2}}{2}\right), \quad (5.40)$$

where the maximum is achieved for the uniform distribution ($\pi_0 = \pi_1 = \frac{1}{2}$) and for the sharp observable M given by the orthonormal basis situated symmetrically with respect to the vectors $|\psi_0\rangle, |\psi_1\rangle$, see Levitin [144]. Here, $\epsilon = |\langle\psi_0|\psi_1\rangle| = \cos \alpha$, where α is the angle between the vectors. Note that in this case the optimal observable is the same as the one obtained by the maximum likelihood criterion (Example 2.26), while C_1 is just the capacity (4.24) of the classical binary symmetric channel, with error probability

$$p = \sin^2(\pi/4 - \alpha/2) = \frac{1 - \sin \alpha}{2}.$$

The plot of the quantities C_1, C as functions of ϵ is given in Figure 5.1.

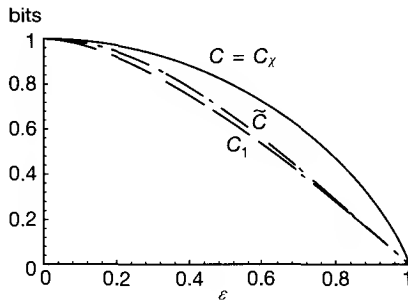


Figure 5.1. The capacity of the binary pure-state channel.

5.5 Typical projectors

Before proving the direct statement of the coding theorem for the quantity C_χ , we shall consider the important notion of the typical subspace due to Schumacher and Jozsa [125].

Consider the tensor degree of a density operator

$$S^{\otimes n} = S \otimes \cdots \otimes S$$

in the space $\mathcal{H}^{\otimes n}$. Let

$$S = \sum_j \lambda_j |e_j\rangle\langle e_j|,$$

be the spectral decomposition of the operator S , where $\{e_j\}$ is an orthonormal basis of eigenvectors and λ_j are the corresponding eigenvalues. Note that λ_j form a probability distribution with entropy

$$\sum_j \eta(\lambda_j) = H(S).$$

Then the spectral decomposition of the product state

$$S^{\otimes n} = S \otimes \cdots \otimes S$$

can be written as

$$S^{\otimes n} = \sum_J \lambda_J |e_J\rangle \langle e_J|,$$

where $J = (j_1, \dots, j_n)$, with eigenvectors $|e_J\rangle = |e_{j_1}\rangle \otimes \cdots \otimes |e_{j_n}\rangle$ and eigenvalues $\lambda_J = \lambda_{j_1} \cdots \lambda_{j_n}$, describing a probability distribution of i.i.d. classical random variables. In accordance with Definition 4.2, the eigenvector $|e_J\rangle$ will be called δ -typical if

$$2^{-n(H(S)+\delta)} < \lambda_J < 2^{-n(H(S)-\delta)}.$$

From the spectral decomposition of $S^{\otimes n}$ we can now construct the projector $P^{n,\delta}$ onto the *typical subspace* $\mathcal{H}^{n,\delta} = P^{n,\delta} \mathcal{H}^{\otimes n}$ by the relation

$$P^{n,\delta} = \sum_{J \in T^{n,\delta}} |e_J\rangle \langle e_J|.$$

We also call $P^{n,\delta}$ the *typical projector*.

The properties of the typical subspace $\mathcal{H}^{n,\delta}$ are analogous to the properties of the set $T^{n,\delta}$ of typical words (see Section 4.1) and easily follow from the latter:

- i. The dimensionality $\dim \mathcal{H}^{n,\delta} = \text{Tr } P^{n,\delta} = |T^{n,\delta}|$ satisfies the inequality

$$\dim \mathcal{H}^{n,\delta} \leq 2^{n(H(S)+\delta)}.$$

- ii. The contribution of the non δ -typical eigenvectors to the operator $S^{\otimes n}$ can be made arbitrarily small, i.e. for given $\delta, \varepsilon > 0$ and large enough n

$$\text{Tr } S^{\otimes n} (I - P^{n,\delta}) < \varepsilon. \quad (5.41)$$

To see this, we evaluate the trace in the basis of the eigenvectors of the operator $S^{\otimes n}$, and find that it is equal to

$$\begin{aligned} \sum_{J \in T^{n,\delta}} \lambda_J &= \text{P} \left\{ \left| -\frac{1}{n} \log \lambda_J - H(S) \right| \geq \delta \right\} \\ &= \text{P} \left\{ \left| -\frac{1}{n} \sum_{k=1}^n \log \lambda_{j_k} - H(S) \right| \geq \delta \right\}, \end{aligned} \quad (5.42)$$

where P corresponds to the probability distribution $\{\lambda_J\}$ and, by the Law of Large Numbers, the last probability can be made arbitrarily small for n large.

iii. For sufficiently large n

$$(1 - \varepsilon)2^{n(H(S) - \delta)} \leq \dim \mathcal{H}^{n, \delta}.$$

Now, we can formulate the noncommutative version of the *asymptotic equipartition property*: for large n the state $S^{\otimes n}$ approaches the chaotic state in the subspace $\mathcal{H}^{n, \delta}$ of dimensionality $\approx 2^{nH(S)}$.

An immediate application of this property is the quantum analog of data compression. As we have seen, the dimensionality of the underlying Hilbert space reflects the potential classical information resource of the quantum system. On the other hand, e.g. in quantum computation, the logarithm of the dimensionality measures the size of the memory, i.e. the number of qubits in the register, which is the most important characteristic that one would like to minimize. Consider the problem of encoding quantum states into other states in a Hilbert space of the smallest possible dimensionality without essential loss of the “quantum information” carried by the states. To this end, we consider a quantum source that produces the *pure* states $S_x = |\psi_x\rangle\langle\psi_x|$ with probabilities p_x . The words $w = (x_1, \dots, x_n)$ generate the product states $S_w = |\psi_w\rangle\langle\psi_w|$, where $|\psi_w\rangle = |\psi_{x_1}\rangle \otimes \dots \otimes |\psi_{x_n}\rangle \in \mathcal{H}^{\otimes n}$, which are then encoded in some other states S'_w in a subspace $\mathcal{H}_d \subset \mathcal{H}^{\otimes n}$, the dimensionality of which $d = \dim \mathcal{H}_d$ should be as small as possible. The fidelity of the encoding is measured by the quantities $\langle\psi_w|S'_w|\psi_w\rangle$. The closer they are to 1, the better the encoding. We require that the *average fidelity*

$$F_n = \sum_w \pi_w \langle\psi_w|S'_w|\psi_w\rangle$$

converges to 1 as $n \rightarrow \infty$.

Let

$$\bar{S}_\pi = \sum_{x=1}^d \pi_x |\psi_x\rangle\langle\psi_x|$$

be the average state of the source.

The following is called the Quantum Data Compression Theorem.

Theorem 5.22 (Schumacher and Jozsa [125]).

- i. For all small enough $\varepsilon, \delta > 0$ there exist a Hilbert space $\mathcal{H}_d \subset \mathcal{H}^{\otimes n}$ of dimensionality $d = 2^{n(H(\bar{S}_\pi) + \delta)}$ and density operators S'_w in \mathcal{H}_d , such that the fidelity $F_n > 1 - \varepsilon$ for n large enough.
- ii. For any choice of the subspace $\mathcal{H}_d \subset \mathcal{H}^{\otimes n}$ with $d = 2^{n(H(\bar{S}_\pi) - \delta)}$ and of the density operators S'_w in \mathcal{H}_d it holds that $F_n < \varepsilon$ for n large enough.

Proof.

- i. Consider the source state $|\psi_w\rangle = |\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_n}\rangle$. We now define the “compressed” state as

$$S'_w = \frac{P^{n,\delta} |\psi_w\rangle \langle \psi_w| P^{n,\delta}}{\langle \psi_w | P^{n,\delta} | \psi_w \rangle}, \quad (5.43)$$

where $P^{n,\delta}$ is the typical projector of the state $\bar{S}_\pi^{\otimes n}$. The operator S'_w acts in the typical subspace $\mathcal{H}^{n,\delta}$, which has a dimensionality of at most $d = 2^{n(H(\bar{S}_\pi) + \delta)}$. For the average fidelity we obtain

$$\begin{aligned} F_n &= \sum_w \pi_w \langle \psi_w | S'_w | \psi_w \rangle \\ &= \sum_w \pi_w \langle \psi_w | P^{n,\delta} | \psi_w \rangle = \text{Tr } \bar{S}_\pi^{\otimes n} P^{n,\delta}, \end{aligned}$$

because $\bar{S}_\pi^{\otimes n} = \sum \pi_w |\psi_w\rangle \langle \psi_w|$. Due to the second property of $P^{n,\delta}$, we can bound this quantity from below by $1 - \varepsilon$.

- ii. Taking into account the fact that the inequality $S'_w \leq P_d$ holds for all S'_w in $\mathcal{H}_d = P_d \mathcal{H}^{\otimes n}$, we have

$$\begin{aligned} F_n &= \sum \pi_w \langle \psi_w | S'_w | \psi_w \rangle \leq \text{Tr } \bar{S}_\pi^{\otimes n} P_d \\ &= \text{Tr } \bar{S}_\pi^{\otimes n} P^{\frac{\delta}{2},n} P_d + \text{Tr } \bar{S}_\pi^{\otimes n} (I - P^{\frac{\delta}{2},n}) P_d \\ &\leq \|\bar{S}_\pi^{\otimes n} P^{\frac{\delta}{2},n}\| \cdot \text{Tr } P_d + \text{Tr } \bar{S}_\pi^{\otimes n} (I - P^{\frac{\delta}{2},n}). \end{aligned}$$

Here we also used the following general property of the trace: if $T \geq 0$, $\text{Tr } TX \leq \text{Tr } T \|X\|$, which follows from (1.18). By the definition of typical eigenvectors, the corresponding eigenvalues are upper-bounded by

$$\|\bar{S}_\pi^{\otimes n} P^{n,\delta}\| < 2^{-n(H(\bar{S}_\pi) - \delta)}. \quad (5.44)$$

By using this inequality and the property ii. of the typical projector, we obtain that for n large enough the quantity F_n does not exceed

$$2^{-n(H(\bar{S}_\pi) - \delta/2)} d + \varepsilon.$$

Thus, if we take $d = 2^{n(H(\bar{S}_\pi) - \delta)}$, $F_n \leq 2^{-n\delta/2} + \varepsilon \leq 2\varepsilon$ for n large enough. \square

For the proof of the direct statement of the Coding Theorem in the next section, we will also need the notion of a conditionally typical projector. Consider the product

states $S_w = S_{x_1} \otimes \cdots \otimes S_{x_n}$, where S_x is now an arbitrary (not necessarily pure) state and $w = (x_1, \dots, x_n)$ is a word in the input alphabet. Denote by

$$\bar{H}_\pi(S_{(\cdot)}) = \sum_x \pi_x H(S_x)$$

the quantum analog of the conditional entropy of the output with respect to the input. Let $P_w = P_w^{n,\delta}$ be the spectral projector of S_w corresponding to the eigenvalues in the interval $(2^{-n[\bar{H}_\pi(S_{(\cdot)})+\delta]}, 2^{-n[\bar{H}_\pi(S_{(\cdot)})-\delta]})$. In more detail, consider the spectral decomposition

$$S_x = \sum_j \lambda_j^x |e_j^x\rangle \langle e_j^x|.$$

In this case, the spectral decomposition of the operator S_w has the form

$$S_w = \sum_J \lambda_J^w |e_J^w\rangle \langle e_J^w|,$$

where $\lambda_J^w = \lambda_{j_1}^{x_1} \cdots \lambda_{j_n}^{x_n}$ are the eigenvalues and $|e_J^w\rangle = |e_{j_1}^{x_1}\rangle \otimes \cdots \otimes |e_{j_n}^{x_n}\rangle$ are the corresponding eigenvectors. The *conditionally typical projector* is defined as

$$P_w = \sum_{J \in T_w^{n,\delta}} |e_J^w\rangle \langle e_J^w|,$$

where

$$T_w^{n,\delta} = \left\{ J : 2^{-n[\bar{H}_\pi(S_{(\cdot)})+\delta]} < \lambda_J^w < 2^{-n[\bar{H}_\pi(S_{(\cdot)})-\delta]} \right\}.$$

The essential properties of the projector P_w are:

i. from the definition,

$$P_w \leq S_w 2^{n[\bar{H}_\pi(S_{(\cdot)})+\delta]}, \quad (5.45)$$

ii. for $\varepsilon > 0$ and sufficiently large n

$$\mathbb{E} \text{Tr } S_w (I - P_w) \leq \varepsilon, \quad (5.46)$$

where \mathbb{E} is the expectation corresponding to the probability distribution

$$\mathbb{P}\{w = (x_1, \dots, x_n)\} = \pi_{x_1} \cdots \pi_{x_n}; \quad (5.47)$$

To prove the second property, consider the sequence of independent trials with outcomes (x_l, j_l) ; $l = 1, \dots, n$, where the probability of an outcome (x, j) is, in each trial, equal to $\pi_x \lambda_j^x$. In this case,

$$\mathbb{E} \text{Tr } S_w (I - P_w) = \mathbb{P}\left\{ J \in \overline{T_w^{n,\delta}} \right\}, \quad (5.48)$$

which tends to 0 as $n \rightarrow \infty$, according to the Law of Large Numbers, see Exercise 4.15.

5.6 Proof of the Direct Coding Theorem

According to the Definition 5.2, it is sufficient to show that the minimal average error probability $\bar{p}_e(n, 2^{nR})$ tends to zero, as $n \rightarrow \infty$, provided $R < C_X$.

Consider the average state

$$\bar{S}_\pi = \sum_x \pi_x S_x$$

and the typical projector $P = P^{n,\delta}$ of the state $\bar{S}_\pi^{\otimes n}$, defined in the previous section. For a given codebook $W = \{w^{(1)}, \dots, w^{(N)}\}$, we can also define the conditionally typical projectors $P_{w^{(j)}}; j = 1, \dots, N$. We now introduce the special suboptimal observable M . The classical Shannon's Coding Theorem 4.13 corresponds to the case of diagonal density operators S_x with conditional probabilities $p(y|x)$ on the diagonal. The projector P is a quantum analog of the indicator of the subset of all δ -typical output words, while $P_{w^{(j)}}$ is the same for the subset of all conditionally typical words, given the input word $w^{(j)}$. However, in the quantum case these projectors need not commute, which makes a straightforward generalization of the decision domains (4.43) and the corresponding measure-theoretic argument impossible. Therefore, we introduce the following observable M in $\mathcal{H}^{\otimes n}$ by letting

$$M_j = \left(\sum_{l=1}^N P P_{w^{(l)}} P \right)^{-1/2} P P_{w^{(j)}} P \left(\sum_{l=1}^N P P_{w^{(l)}} P \right)^{-1/2} \quad ; \quad j = 1, \dots, N. \quad (5.49)$$

The normalization with the square root is necessary, because the sum of the operators $P P_{w^{(j)}} P$ may not be equal to the unit operator, and the replacement of $P_{w^{(j)}}$ by $P P_{w^{(j)}} P$ plays the role of intersection with the set of all typical words in the formula (4.44). The operator $(\sum_{l=1}^N P P_{w^{(l)}} P)^{-1/2}$ is to be understood as the generalized inverse of $(\sum_{l=1}^N P P_{w^{(l)}} P)^{1/2}$, equal to 0 on the null subspace of that operator, which contains the range of the projector $I - P$. Denoting by \hat{P} the projector onto the support of $\sum_{l=1}^N P P_{w^{(l)}} P$, we have

$$P P_{w^{(l)}} P \leq \hat{P} \leq P, \quad l = 1, \dots, N. \quad (5.50)$$

To avoid cumbersome notations, we shall further enumerate words by the variable w , omitting the indices j, l . By denoting

$$A_w = P_w P \left(\sum_{w'=1}^N P P_{w'} P \right)^{-1/2}$$

and using the Cauchy–Schwarz inequality (2.24),

$$|\text{Tr} S_w A_w|^2 \leq \text{Tr} S_w A_w^* A_w,$$

we obtain

$$\bar{P}_e(W, M) \leq \frac{1}{N} \sum_{w=1}^N [1 - |\text{Tr} S_w A_w|^2] \leq \frac{2}{N} \sum_{w=1}^N [1 - \text{Tr} S_w A_w],$$

where $\text{Tr} S_w A_w = \text{Tr} P S_w P_w P (\sum_{w'=1}^N P P_{w'} P)^{-1/2}$ is a real number between 0 and 1 (note that $S_w P_w = P_w S_w \geq 0$). Applying the inequality

$$-2x^{-1/2} \leq -3 + x, \quad x > 0, \quad (5.51)$$

we obtain by (5.50)

$$-2 \left(\sum_{w'=1}^N P P_{w'} P \right)^{-1/2} \leq -3 \hat{P} + \sum_{w'=1}^N P P_{w'} P \leq -3 P P_w P + \sum_{w'=1}^N P P_{w'} P.$$

Therefore, by using the inequality (1.13),

$$\begin{aligned} \bar{P}_e(W, M) &\leq \frac{1}{N} \sum_{w=1}^N [2\text{Tr} S_w - 3\text{Tr} S_w P_w P P_w P + \sum_{w'=1}^N \text{Tr} S_w P_w P P_{w'} P] \\ &= \frac{1}{N} \sum_{w=1}^N [2\text{Tr} S_w (I - P_w P P_w P) + \sum_{w': w' \neq w} \text{Tr} S_w P_w P P_{w'} P]. \end{aligned}$$

Taking into account that

$$\begin{aligned} \text{Tr} S_w (I - P_w P P_w P) &= \text{Tr} S_w (I - P_w) P P_w P + \text{Tr} S_w (I - P) P_w \\ &\quad - \text{Tr} S_w (I - P) P_w (I - P) + \text{Tr} S_w (I - P_w) P \\ &\quad + \text{Tr} S_w (I - P) \\ &\leq 2[\text{Tr} S_w (I - P_w) + \text{Tr} S_w (I - P)], \end{aligned}$$

where we used (1.13), we can write

$$\bar{P}_e(W, M) \leq \frac{1}{N} \sum_{w=1}^N [4\text{Tr} S_w (I - P) + 4\text{Tr} S_w (I - P_w) + \sum_{w': w' \neq w} \text{Tr} P S_w P P_{w'}], \quad (5.52)$$

which is our final basic estimate, similar to (4.45) in the classical case.

We now again apply Shannon's random coding scheme, assuming that the words $w^{(1)}, \dots, w^{(N)}$ are chosen at random, independently, and with the probability distribution (5.47) for each word, where π is the optimal distribution for which

$$H(\bar{S}_\pi) - \bar{H}_\pi(S_{(\cdot)}) = C_\chi.$$

Then

$$\mathbb{E}S_w = \sum_{x_1 \dots x_n} \pi_{x_1} \dots \pi_{x_n} S_{x_1} \otimes \dots \otimes S_{x_n} = \bar{S}_\pi^{\otimes n}.$$

Taking the expectations in (5.52) and using the independence of random operators $S_w, P_{w'}$, we obtain

$$\begin{aligned} \mathbb{E}\bar{P}_e(W, M) &\leq 4\mathrm{Tr} \bar{S}_\pi^{\otimes n} (I - P) + 4\mathbb{E}\mathrm{Tr} S_w (I - P_w) \\ &\quad + (N - 1)\mathrm{Tr} \bar{S}_\pi^{\otimes n} P \mathbb{E}P_{w'}. \end{aligned} \quad (5.53)$$

By inequalities (5.41), (5.46), expressing typicality of the projectors P, P_w , and by inequality (1.18) we have

$$\mathbb{E}\bar{P}_e(W, M) \leq 8\epsilon + (N - 1)\|\bar{S}_\pi^{\otimes n} P\| \mathbb{E}P_{w'},$$

for $n \geq n(\pi, \epsilon, \delta)$. By the property (5.44) of the projector P ,

$$\|\bar{S}_\pi^{\otimes n} P\| \leq 2^{-n[H(\bar{S}_\pi) - \delta]},$$

and by the property i. of the projector P_w ,

$$\mathbb{E}P_{w'} = \mathbb{E}\mathrm{Tr} P_{w'} \leq \mathbb{E}\mathrm{Tr} S_{w'} \cdot 2^{n[\bar{H}_\pi(S_{(\cdot)}) + \delta]} = 2^{n[\bar{H}_\pi(S_{(\cdot)}) + \delta]}.$$

Remembering that $H(\bar{S}_\pi) - \bar{H}_\pi(S_{(\cdot)}) = C_\chi$, we obtain

$$\bar{p}_e(n, N) \leq \mathbb{E}\bar{P}_e(W, M) \leq 8\epsilon + N 2^{-n[C_\chi - 2\delta]}.$$

Thus, if $R \leq C_\chi - 3\delta$,

$$\bar{p}_e(n, 2^{nR}) \leq 8\epsilon + 2^{-n\delta},$$

hence, the Direct Coding Theorem follows. \square

A slight modification of this argument shows exponential decay of the average error probability. Namely, there exists a $\beta > 0$, such that

$$\mathbb{E}\bar{P}_e(W, M) \leq 2^{-n\beta}. \quad (5.54)$$

Indeed, we already have such an estimate for the last term in (5.53), provided $N \leq 2^{n[C_\chi - 3\delta]}$, so we have to establish it for the first two terms. But this follows by application of inequality (4.11) to the probabilities of large deviations (5.42), (5.48) (**Exercise**).

5.7 The reliability function for pure-state channel

In the classical case, pure signal states correspond to degenerate distributions, and the Coding Theorem trivially implies the maximal value of the channel capacity $\log |\mathcal{X}|$. However, for pure quantum nonorthogonal states, the Coding Theorem remains non-trivial, although the situation is simplified, allowing us to obtain a stronger result. Here, we give a different proof of the direct statement of the Coding Theorem in the case of pure states $S_x = |\psi_x\rangle\langle\psi_x|$, which does not use the notion of typicality but, due to a more sophisticated analysis, gives the estimate for the exponential speed of decay of the error probability.

Theorem 5.23 (Burnashev and Holevo [33]). *For $R < C_X$, the following inequality holds*

$$\bar{p}_e(n, 2^{nR}) \leq 2 \cdot 2^{-nE(R)},$$

where

$$E(R) = \max_{0 \leq r \leq 1} \left[-Rr + \max_{\pi} (-\log \text{Tr } \bar{S}_{\pi}^{1+r}) \right] > 0. \quad (5.55)$$

The function $E(R)$ gives a lower bound for the so called *reliability function* of the channel, characterizing the exponential speed of decay of the error probability $\bar{p}_e(n, 2^{nR})$ as $n \rightarrow \infty$.

Proof. Consider the subspace of the space $\mathcal{H}^{\otimes n}$, spanned by the code vectors $\psi_{w(1)}, \dots, \psi_{w(N)}$. As shown in Section 2.3.4, we can use the Gram operator $G = \sum_j |\psi_{w(j)}\rangle\langle\psi_{w(j)}|$ to construct the overcomplete system

$$|\hat{\psi}_j\rangle = G^{-\frac{1}{2}} |\psi_{w(j)}\rangle, \quad j = 1, \dots, N,$$

and hence the observable M with the components

$$M_j = |\hat{\psi}_j\rangle\langle\hat{\psi}_j|, \quad j = 1, \dots, N, \quad (5.56)$$

which can be extended to the whole space $\mathcal{H}^{\otimes n}$ by adjoining the projector M_0 onto the orthogonal complement to $\psi_{w(1)}, \dots, \psi_{w(N)}$. This will be our suboptimal decoding for the composite channel. Note that it differs from (5.49), in that it does not contain the typical projector P (while the conditionally typical projectors in the case of pure states are $P_w = |\psi_w\rangle\langle\psi_w|$). In this way, we obtain the upper bound for the average error probability expressed via the Gram operator of the code vectors:

$$\bar{P}_e(W, M) = \frac{1}{N} \sum_{j=1}^N \left(1 - \langle \hat{\psi}_j | G^{\frac{1}{2}} | \hat{\psi}_j \rangle^2 \right)$$

$$\begin{aligned}
&\leq \frac{2}{N} \sum_{j=1}^N \left(1 - \langle \hat{\psi}_j | G^{\frac{1}{2}} | \hat{\psi}_j \rangle\right) \\
&= \frac{2}{N} \left(N - \text{Tr } G^{\frac{1}{2}}\right)
\end{aligned} \tag{5.57}$$

Assuming that the codewords $\{w^{(j)}\}$ are randomly chosen, independently of each other, with the distribution (5.47), we obtain

$$\mathbb{E} \bar{P}_e(W, M) \leq \frac{2}{N} \left(N - \mathbb{E} \text{Tr } G^{\frac{1}{2}}\right). \tag{5.58}$$

By using the inequality

$$-2x^{\frac{1}{2}} \leq x^2 - 3x, \quad x \geq 0,$$

which follows from (5.51), in combination with the obvious one $-2x^{\frac{1}{2}} \leq 0$, we have

$$-2G^{\frac{1}{2}} \leq \begin{cases} G^2 - 3G \\ 0 \end{cases}. \tag{5.59}$$

The expectation of the Gram operator is equal to

$$\begin{aligned}
\mathbb{E} G &= \mathbb{E} \sum_{j=1}^N |\psi_{w^{(j)}}\rangle \langle \psi_{w^{(j)}}| = N \mathbb{E} |\psi_{w^{(j)}}\rangle \langle \psi_{w^{(j)}}| \\
&= N \sum_{x_1 \dots x_n} \pi_{x_1} \dots \pi_{x_n} (|\psi_{x_1}\rangle \otimes \dots \otimes |\psi_{x_n}\rangle) (\langle \psi_{x_1}| \otimes \dots \otimes \langle \psi_{x_n}|) \\
&= N \left[\sum_x \pi_x |\psi_x\rangle \langle \psi_x| \right]^{\otimes n} = N \bar{S}_\pi^{\otimes n}.
\end{aligned} \tag{5.60}$$

Similarly

$$\begin{aligned}
\mathbb{E} (G^2 - G) &= \mathbb{E} \left(\sum_{j,k=1}^N |\psi_{w^{(j)}}\rangle \langle \psi_{w^{(j)}}| \psi_{w^{(k)}}\rangle \langle \psi_{w^{(k)}}| \right. \\
&\quad \left. - \sum_{j=1}^N |\psi_{w^{(j)}}\rangle \langle \psi_{w^{(j)}}| \right) \\
&= \mathbb{E} \left(\sum_{j \neq k} |\psi_{w^{(j)}}\rangle \langle \psi_{w^{(j)}}| \psi_{w^{(k)}}\rangle \langle \psi_{w^{(k)}}| \right) \\
&= N(N-1) \left(\bar{S}_\pi^{\otimes n} \right)^2,
\end{aligned} \tag{5.61}$$

so that in combination with (5.59) we obtain

$$-2EG^{\frac{1}{2}} \leq \begin{cases} N(N-1)(\bar{S}_\pi^{\otimes n})^2 - 2N\bar{S}_\pi^{\otimes n} \\ 0 \end{cases}. \quad (5.62)$$

Let $\{|e_J\rangle\}$ be the orthonormal basis of eigenvectors and $\{\lambda_J\}$ the corresponding eigenvalues of the operator $\bar{S}_\pi^{\otimes n}$. In this case,

$$-2\langle e_J | EG^{-\frac{1}{2}} | e_J \rangle \leq \begin{cases} N(N-1)\lambda_J^2 - 2N\lambda_J \\ 0 \end{cases}, \quad (5.63)$$

hence, using the inequalities $\min(a, b) \leq a^r b^{1-r}$ and $2^{1-r} \leq 2$ for $0 \leq r \leq 1$, we obtain

$$N\lambda_J \min[(N-1)\lambda_J, 2] - 2N\lambda_J \leq 2N(N-1)^r \lambda_J^{1+r} - 2N\lambda_J, \quad (5.64)$$

whence

$$-2E\text{Tr } G^{\frac{1}{2}} \leq -2N + 2N(N-1)^r \text{Tr}(\bar{S}_\pi^{\otimes n})^{1+r}.$$

Coming back to (5.58), we can evaluate the average error probability minimized over all codes as

$$\bar{p}_e(n, N) \leq 2N^r (\text{Tr } \bar{S}_\pi^{1+r})^n$$

for any $0 \leq r \leq 1$. Putting $N = 2^{nR}$, we obtain

$$\bar{p}_e(n, 2^{nR}) \leq 2 \cdot 2^{n(Rr + \log \text{Tr } \bar{S}_\pi^{1+r})}. \quad (5.65)$$

$$(5.66)$$

Now, we can use the freedom in the choice of the parameters r and π . Note that $-\log \text{Tr } \bar{S}_\pi^{1+r}$ is a concave function of r , which can be established by showing that the second derivative is negative (**Exercise**). Also,

$$\left. \frac{d}{dr} (-\log \text{Tr } \bar{S}_\pi^{1+r}) \right|_{r=0} = -\text{Tr } \bar{S}_\pi \log \bar{S}_\pi = H(\bar{S}_\pi). \quad (5.67)$$

Take for π the optimal distribution, for which $H(\bar{S}_\pi) = C_\chi$. Then, for $R < C_\chi$ we have $\bar{p}_e(n, 2^{nR}) \leq 2 \cdot 2^{-nE(R)}$, where $E(R)$ is given by expression (5.55). \square

5.8 Notes and references

1. The issue of the information capacity of quantum communication channels arose in the sixties in the papers of Gordon [71], Forney [60], Lebedev and Levitin [142], and Stratonovich [203] and goes back to even earlier, classical works of Gabor and Brillouin, that ask for fundamental physical limits on the rate and quality of information transmission (for detailed reference, see the surveys [38], [98] and the books [86], [107]). These works laid a physical foundation and raised the question of consistent quantum information treatment of the problem.

2. The proof of Theorem 5.9, including the criterion for the equality, based on comparison of the convexity of the quantities in (5.16), was given in [95]. The symmetric logarithmic derivative of a family of quantum states and the corresponding quantum Fisher information were introduced by Helstrom in the context of quantum estimation theory [86].

3. A detailed comparison of the capacities C_1 , C_χ for different channels was made in the paper of Hirota et al. [127].

The Strong Converse

$$\lim_{n \rightarrow \infty} \bar{p}_e(n, 2^{nR}) = 1 \quad \text{for } R \geq C_\chi$$

was obtained by Ogawa and Nagaoka [160], see also the book [78].

Formula (5.40) follows from the work of Levitin [144], who showed that for two pure states and arbitrary π the maximum $\max_M \mathcal{J}(\pi, M)$ over the two-valued sharp observables in the plane is attained by the observable that maximizes the average probability of correct decision, and from the work of Shor [192], who proved that the two-valued sharp observables are sufficient in this case, and the maximum $\max_M \mathcal{J}(\pi, M)$ is a concave symmetric function of π , implying that the optimal distribution is uniform.

The trine channel was introduced by Holevo [94], who proved that for the uniform distribution π , the maximum $\max_M \mathcal{J}(\pi, M)$ over all observables is strictly greater than the maximum over sharp observables ($\log(\sqrt{3}/\sqrt[3]{2}) \approx 0.459$). It is interesting that the information-optimal observable in this problem differs from the maximal likelihood observable (see Section 2.4.2) by rotation by the angle $\pi/2$, the latter being the least informative observable. A complete treatment of $\max_M \mathcal{J}(\pi, M)$ for an arbitrary, symmetric family of pure states in two-dimensional Hilbert space was given by Sasaki, Barnett, Jozsa, Osaki and Hirota [176], who also discussed experimental implementation of the optimal measuring procedures. Conjecture (5.39) was confirmed numerically in [126], and an additional argument was given in [192]. A remarkable study of the “lifted trine” channel, i.e. the channel with three linearly independent, equiangular state vectors in 3-dimensional Hilbert space was given by Shor [192]. It

was shown, in particular, that the information-optimal observable in this case has in general 6 outcomes, i.e. the maximal value admitted by Proposition 5.8 in the case $d = 3$ (although the gain, as compared to 3 outcomes, is tiny).

4. The idea of quantum data compression, originally due to Schumacher, is one of the central ones in quantum information theory. Theorem 5.22 is proved by Jozsa and Schumacher [125]. For a survey of more recent results, from different authors, on the much more complicated problem of data compression for sources with mixed states, see the book of Hayashi [78].

5. The proof of the coding theorem for a general c-q channel was given by Holevo [97] and by Schumacher and Westmoreland [177]. See Winter [223] for a different proof in the spirit of Csiszár and Körner's approach [44] to the classical coding theorem. This was preceded by the consideration of the pure-state channel in the work of Hausladen, Jozsa, Schumacher, Westmoreland and Wootters [77], who used the technique of typical subspaces discussed in Section 5.5. It should be noted that Shannon's method of random coding for the channel with pure, almost orthogonal states was first applied by Stratonovich and Vancian [205], who used the first term in the minimal error probability in the asymptotics $\varepsilon \rightarrow 1$ (almost orthogonal states) to conclude that this method allows us to obtain the so called cut-off rate

$$\tilde{C} = -\log \min_{\pi} \text{Tr} \bar{\rho}_{\pi}^2 < C. \quad (5.68)$$

As shown by Burnashev and Holevo [33], the properly used method of random coding allows us to obtain not only the classical capacity C , but also the exponential estimate for the error probability, in the spirit of the classical results in Gallager [66]. Recently, Dalai [45] was able to extend the classical sphere packing bound of Fano, Shannon, Gallager and Berlecamp to the pure state c-q channel, to obtain the upper bound on the reliability function, which coincides, at high rates, with the lower bound $E(R)$ of [33]. The paper [33] also contains a formulation of the hypothesis concerning the reliability function for an arbitrary mixed-state channel. In this case, like in the others, a mixed-state problem is much more complex than its pure-state analog.

Part III

Channels and entropies

Chapter 6

Quantum evolutions and channels

6.1 Quantum evolutions

In this section, the notion of quantum communication channel will be considered from a general point of view. As we have seen in Chapter 4, a classical communication channel is completely characterized by the transition probability matrix

$$\mathcal{X} \xrightarrow{p(y|x)} \mathcal{Y}.$$

Such a channel describes an affine transformation $p_x \rightarrow p'_y = \sum_x p(y|x)p_x$ of classical states (probability distributions) $p = \{p_x\}$ on the input alphabet \mathcal{X} , into the classical states $p' = \{p'_y\}$ on the output alphabet \mathcal{Y} .

To have a proper quantum analog of a channel we will therefore first look for maps taking density operators into density operators, which respect convex combinations, i.e. affine maps $\Phi : \mathfrak{S}(\mathcal{H}) \rightarrow \mathfrak{S}(\mathcal{H})$,

$$\Phi \left[\sum_j p_j S_j \right] = \sum_j p_j \Phi[S_j]; \quad p_j \geq 0, \quad \sum_j p_j = 1; \quad S_j \in \mathfrak{S}(\mathcal{H}).$$

Exercise 6.1. Prove the following statement: an affine map Φ of $\mathfrak{S}(\mathcal{H})$ into itself can be uniquely extended to a map of the linear space $\mathfrak{T}(\mathcal{H})$ with the properties

- i. Φ is linear: $\Phi \left[\sum_j c_j S_j \right] = \sum_j c_j \Phi[S_j]$, $c_j \in \mathbb{C}$, $S_j \in \mathfrak{T}(\mathcal{H})$
- ii. Φ is positive: $S \in \mathfrak{T}(\mathcal{H})$, $S \geq 0 \Rightarrow \Phi[S] \geq 0$
- iii. Φ is trace preserving: $\text{Tr } \Phi[S] = \text{Tr } S$, $S \in \mathfrak{T}(\mathcal{H})$

Hint: Follow the proof of Theorem 2.6 to obtain the linear extension of Φ to $\mathfrak{T}(\mathcal{H})$. Properties ii., iii. then are straightforward.

Definition 6.2. For every map $\Phi : \mathfrak{T}(\mathcal{H}) \rightarrow \mathfrak{T}(\mathcal{H})$, satisfying the properties i.-iii., the dual map $\Phi^* : \mathfrak{B}(\mathcal{H}) \rightarrow \mathfrak{B}(\mathcal{H})$ is defined by the formula

$$\text{Tr } \Phi[S]X = \text{Tr } S\Phi^*[X], \quad S \in \mathfrak{T}(\mathcal{H}), \quad X \in \mathfrak{B}(\mathcal{H}). \quad (6.1)$$

Exercise 6.3. Prove the following statement: the properties of the map Φ from Exercise 6.1 are equivalent to the following:

- i. Φ^* is linear;
- ii. Φ^* is positive: $X \geq 0 \Rightarrow \Phi^*[X] \geq 0$;
- iii. Φ^* is unital: $\Phi^*[I] = I$.

The map Φ describes an evolution of the quantum system in terms of states (the *Schrödinger picture*), while the map Φ^* does this in terms of observables (the *Heisenberg picture*). Notice that positive linear maps transform Hermitian operators into Hermitian operators. In the case of a finite-dimensional \mathcal{H} that we are considering, the linear space $\mathfrak{T}(\mathcal{H})$ and the algebra $\mathfrak{B}(\mathcal{H})$ coincide with the space of all linear operators in \mathcal{H} , but we use different notations to stress that $\mathfrak{T}(\mathcal{H})$ is the arena for the states and the Schrödinger picture, while $\mathfrak{B}(\mathcal{H})$ is that of the observables and the Heisenberg picture. This distinction becomes quite substantial in the infinite-dimensional case considered in Chapters 11, 12.

Example 6.4. Let U be a unitary operator. In this case, $\Phi[S] = USU^*$ is an affine, one-to-one mapping of the set of quantum states $\mathfrak{S}(\mathcal{H})$ onto itself (i.e. affine bijection), describing *reversible* evolution. In the Heisenberg picture $\Phi^*[X] = U^*XU$.

The following result, going back to the famous *Wigner's Theorem*, characterizes all reversible evolutions.

Theorem 6.5. Let Φ be an affine bijection of the convex set $\mathfrak{S}(\mathcal{H})$. In this case,

$$\Phi[S] = USU^*, \quad \text{or} \quad \Phi[S] = US^\top U^*, \quad (6.2)$$

where U is unitary operator and S^\top is the matrix transposition in some basis. Equivalently, $\Phi[S] = USU^*$, where U is a unitary or antiunitary operator.

An antiunitary operator U is characterized by the properties

- i. $\|U\psi\| = \|\psi\|; \psi \in \mathcal{H}$;
- ii. $U(\sum c_j \psi_j) = \sum \bar{c}_j U\psi_j$.

Such an operator can always be represented in the form $U = \tilde{U}\Lambda$, where \tilde{U} is a unitary operator and $\Lambda = \Lambda^*$ is the antiunitary operator of the complex conjugation, in a fixed basis. The corresponding evolution of states is given by the matrix transposition in this basis

$$S^\top = \Lambda S \Lambda^*.$$

In physics, transposition and complex conjugation are associated with time inversion.

Proof. We rely on the case of the qubit ($\dim \mathcal{H} = 2$), for which the proof will be given in Section 6.8. Let ψ_1, ψ_2 be some linearly independent vectors in \mathcal{H} . Since an affine bijection of a convex set maps extreme points into extreme points, we have

$$\Phi[|\psi_j\rangle\langle\psi_j|] = |\phi_j\rangle\langle\phi_j|; \quad j = 1, 2. \quad (6.3)$$

Denote by \mathcal{H}_2 the two-dimensional subspace generated by ψ_1, ψ_2 . In this case, for any $|\psi\rangle \in \mathcal{H}_2$ we have

$$|\psi\rangle\langle\psi| \leq c (|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|).$$

On the other hand, similar to (6.3),

$$\Phi[|\psi\rangle\langle\psi|] = |\phi\rangle\langle\phi|.$$

Because of the positivity of Φ

$$|\phi\rangle\langle\phi| \leq c (|\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2|),$$

i.e. $|\phi\rangle$ belongs to a two-dimensional subspace that contains the vectors ϕ_1, ϕ_2 . By an additional unitary transformation, we can make this subspace coincide with \mathcal{H}_2 . Hence, we can apply the aforementioned result for the qubit to the restriction of Φ onto the subset of states supported by the two-dimensional subspace $\mathcal{H}_2 \subset \mathcal{H}$ to conclude that this restriction is given by the formula (6.2), where U may, however, depend on that subspace.

The proof in the general case is obtained by performing appropriate transformations of the subspace \mathcal{H}_2 to show that this dependence reduces to an inessential phase factor. Let $\{e_j; j = 1, \dots, d\}$ be an orthonormal basis in \mathcal{H} . In this case,

$$\Phi[|e_j\rangle\langle e_j|] = |h_j\rangle\langle h_j|.$$

By applying the aforementioned result for \mathcal{H}_2 generated by the vectors e_j, e_k , we obtain that $\{h_j\}$ is an orthonormal basis. Moreover,

$$\Phi[|e_j\rangle\langle e_k|] = z_{jk}|h_j\rangle\langle h_k| \quad \text{or} \quad z_{kj}|h_k\rangle\langle h_j|, \quad (6.4)$$

where $|z_{jk}| = 1$. If $\Phi[|e_1\rangle\langle e_2|] = z_{12}|h_1\rangle\langle h_2|$ but $\Phi[|e_1\rangle\langle e_3|] = z_{31}|h_3\rangle\langle h_1|$ then

$$\Phi[|e_1\rangle\langle e_2 + e_3|] = z_{12}|h_1\rangle\langle h_2| + z_{31}|h_3\rangle\langle h_1|$$

is an operator of rank 2, which contradicts (6.4). Repeating this argument, we obtain that in relation (6.4) one and the same alternative takes place for all j, k .

Consider the first alternative, when

$$\Phi[|e_j\rangle\langle e_k|] = z_{jk}|h_j\rangle\langle h_k|.$$

Now, putting $\psi = e_1 + \cdots + e_d$, we obtain that the operator

$$\Phi[|\psi\rangle\langle\psi|] = \sum_{j,k=1}^d z_{jk} |h_j\rangle\langle h_k|$$

has rank 1, hence $z_{jk} = a_j \bar{a}_k$, where $|a_j| = 1$ for all j . Denoting by U the unitary operator for which $U|e_j\rangle = a_j|h_j\rangle$ we obtain $\Phi[S] = USU^*$. Similarly, in the case of the second alternative we obtain $\Phi[S] = US^T U^*$, where T is a transposition in the basis $\{e_j\}$. \square

6.2 Completely positive maps

Generalizing the previous considerations, we shall consider a linear map Φ , acting from $\mathfrak{T}(\mathcal{H}_A)$ to $\mathfrak{T}(\mathcal{H}_B)$, where $\mathcal{H}_A, \mathcal{H}_B$ are, in general, different spaces. In this case, the dual map Φ^* acts from $\mathfrak{B}(\mathcal{H}_B)$ to $\mathfrak{B}(\mathcal{H}_A)$.

Definition 6.6. The map Φ^* (Φ) is called *completely positive* (CP) if one of the two equivalent conditions hold:

- i. $\Phi^* \otimes \text{Id}_n$ is positive for all $n = 1, 2, \dots$, where Id_n denotes the identity map of the algebra of all $n \times n$ -matrices $\mathfrak{B}(\mathbb{C}^n)$.
- ii. For arbitrary finite sets of vectors $\{\varphi_j\} \subset \mathcal{H}_B, \{\psi_j\} \subset \mathcal{H}_A$ the following inequality holds:

$$\sum_{j,k} \langle \varphi_j | \Phi[|\psi_j\rangle\langle\psi_k|] \varphi_k \rangle \geq 0.$$

To see the equivalence of the above conditions, let us introduce the spaces $\mathcal{H}_i^{(n)}$, which are direct orthogonal sums of n copies of \mathcal{H}_i ; $i = A, B$, where n is the size of given sets of vectors $\{\varphi_j\}, \{\psi_j\}$. As we have seen in Section 3.1.1, $\mathcal{H}_i^{(n)}$ can be identified with $\mathcal{H}_i \otimes \mathbb{C}^n$. In this case, denoting

$$\varphi^{(n)} = \sum_{j=1}^n \oplus \varphi_j, \quad \psi^{(n)} = \sum_{j=1}^n \oplus \psi_j,$$

we have

$$\sum_{j,k} \langle \varphi_j | \Phi[|\psi_j\rangle\langle\psi_k|] \varphi_k \rangle = \langle \psi^{(n)} | (\Phi^* \otimes \text{Id}_n)[|\varphi^{(n)}\rangle\langle\varphi^{(n)}|] | \psi^{(n)} \rangle,$$

from which the implication i. \Rightarrow ii. follows. Conversely, any positive operator $X^{(n)}$ in $\mathcal{H}_B^{(n)} \simeq \mathcal{H}_B \otimes \mathbb{C}^n$ can be represented as a sum of positive rank one operators of the form $|\varphi^{(n)}\rangle\langle\varphi^{(n)}|$. Hence, condition ii., which means positivity of $\Phi^* \otimes \text{Id}_n$ on such operators, implies i.

Exercise 6.7. Prove the following statement: if the map Φ^* satisfies the condition i. for $n = 2$, the *Kadison inequality* holds

$$\Phi^*[X]^* \Phi^*[X] \leq \|\Phi^*[I]\| \Phi^*[X^*X], \quad (6.5)$$

for arbitrary operator X .

Exercise 6.8. Show that for the transposition map $S \rightarrow S^\top$ in some basis, condition i. already breaks for $n = 2$. Hint: consider condition ii. where $\{\psi_j\}$ is the orthonormal basis in which the transposition is made.

The notion of complete positivity was introduced by Stinespring [202] who proved an important result, generalizing Naimark's Theorem. We shall give the proof of *Stinespring's Dilation Theorem* in the special finite-dimensional case.

Theorem 6.9. For every completely positive map $\Phi^*: \mathfrak{B}(\mathcal{H}_B) \rightarrow \mathfrak{B}(\mathcal{H}_A)$ there exist a Hilbert space \mathcal{H}_E and an operator $V: \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ such that

$$\Phi^*[X] = V^*(X \otimes I_E)V, \quad X \in \mathfrak{B}(\mathcal{H}_B). \quad (6.6)$$

Dually,

$$\Phi[S] = \text{Tr}_E V S V^*, \quad S \in \mathfrak{T}(\mathcal{H}_A). \quad (6.7)$$

The map Φ^* is unital, i.e. $\Phi^*[I_B] = I_A$ (the map Φ preserves the trace) if and only if the operator V is isometric.

In relation (6.7) and in what follows we use the abbreviated notation for the partial trace: $\text{Tr}_E = \text{Tr}_{\mathcal{H}_E}$ etc.

Proof. Consider the algebraic tensor product $\mathcal{L} = \mathcal{H}_A \otimes \mathfrak{B}(\mathcal{H}_B)$, generated by the elements $\Psi = \psi \otimes X$, $\psi \in \mathcal{H}_A$, $X \in \mathfrak{B}(\mathcal{H}_B)$. Let us introduce a pre-inner product in \mathcal{L} with the corresponding square of the norm

$$\left\| \sum_j \psi_j \otimes X_j \right\|^2 = \sum_{j,k} \langle \psi_j | \Phi^*[X_j^* X_k] | \psi_k \rangle.$$

This quantity is nonnegative because the block operator $X^{(n)} = [X_j^* X_k]_{j,k=\overline{1,n}}$ in $\mathcal{H}_B^{(n)}$ is positive and the map Φ is CP. Taking the quotient with respect to the subspace \mathcal{L}_0 of zero norm, we obtain the Hilbert space $\mathcal{K} = \mathcal{L}/\mathcal{L}_0$. Define V and π by the relations $V\psi = \psi \otimes I$ and $\pi[Y]\Psi = \pi[Y](\psi \otimes X) = \psi \otimes YX$. It is easy to check that these definitions agree with taking the quotient. Then π is a *-homomorphism of the algebra $\mathfrak{B}(\mathcal{H}_B)$ into $\mathfrak{B}(\mathcal{K})$, i.e. a linear map that preserves the algebraic operations and the involution:

$$\pi[XY] = \pi[X]\pi[Y], \pi[X^*] = \pi[X]^*.$$

The map π is unital. Moreover,

$$\langle \varphi | \Phi^*[X] | \psi \rangle = \langle \varphi \otimes I | \psi \otimes X \rangle = \langle \varphi | V^* \pi[X] V | \psi \rangle, \quad X \in \mathfrak{B}(\mathcal{H}_B),$$

that is

$$\Phi^*[X] = V^* \pi[X] V. \quad (6.8)$$

However, by Lemma 6.10 proved below any *-homomorphism of the algebra $\mathfrak{B}(\mathcal{H}_B)$ is unitary equivalent to the ampliation $\pi[X] = X \otimes I_E$, where I_E is the unit operator in a Hilbert space \mathcal{H}_E , i.e. we can take $\mathcal{K} = \mathcal{H}_B \otimes \mathcal{H}_E$, and the Stinespring representation (6.8) takes the form (6.6).

The dual representation (6.7) follows from (6.6) and from the definition of the dual map (6.1). \square

In this construction, the dimensionality of the space \mathcal{H}_E satisfies

$$\dim \mathcal{H}_E \leq d_A d_B, \quad (6.9)$$

where $d_A = \dim \mathcal{H}_A$, $d_B = \dim \mathcal{H}_B$. Indeed, $\dim \mathcal{H}_E = \dim \mathcal{K} / d_B \leq \dim \mathcal{L} / d_B$ and $\dim \mathcal{L} = d_A d_B^2$. However, there is no a priori restriction and any map of the form (6.6) is completely positive as concatenation of ampliation $X \rightarrow X \otimes I$ and conjugation $X \rightarrow V^* X V$, the complete positivity of which follows from Definition 6.6.

Lemma 6.10. *Let π be a unital *-homomorphism of the algebra $\mathfrak{B}(\mathcal{H})$ into the algebra $\mathfrak{B}(\mathcal{K})$. In this case, there exist Hilbert space \mathcal{H}_E and isometric map U of the space \mathcal{K} onto $\mathcal{H} \otimes \mathcal{H}_E$ such that*

$$\pi[X] = U^* (X \otimes I_E) U, \quad X \in \mathfrak{B}(\mathcal{H}). \quad (6.10)$$

Proof. Choose an orthonormal basis $\{e_j; j = 1, \dots, d\}$ in \mathcal{H} , and consider the matrix units $|e_j\rangle\langle e_k|$ and their images

$$V_{jk} = \pi[|e_j\rangle\langle e_k|] \in \mathfrak{B}(\mathcal{K}).$$

Since π is *-homomorphism, the operators V_{jk} satisfy the same algebraic relations as the matrix units

$$V_{jk} V_{lm} = \delta_{kl} V_{jm}, \quad V_{jk}^* = V_{kj}. \quad (6.11)$$

Consider the subspace $\mathcal{H}_E = V_{11} \mathcal{K} \subseteq \mathcal{K}$ and define $U : \mathcal{K} \rightarrow \mathcal{H} \otimes \mathcal{H}_E$ by the relation

$$U \psi = \sum_j |e_j\rangle \otimes V_{1j} \psi.$$

(Note that $V_{1j} \psi = V_{11} V_{1j} \psi \in \mathcal{H}_E$). In this case,

$$\langle U \psi_1 | (X \otimes I_E) U \psi_2 \rangle = \sum_{j,k} \langle e_j | X e_k \rangle \langle V_{1j} \psi_1 | V_{1k} \psi_2 \rangle. \quad (6.12)$$

By (6.11)

$$\langle V_{1j}\psi_1 | V_{1k}\psi_2 \rangle = \langle \psi_1 | V_{jk}\psi_2 \rangle = \langle \psi_1 | \pi[|e_j\rangle\langle e_k|]\psi_2 \rangle,$$

Therefore, the right hand side of (6.12) is equal to

$$\sum_{j,k} \langle e_j | X e_k \rangle \langle \psi_1 | \pi[|e_j\rangle\langle e_k|]\psi_2 \rangle = \langle \psi_1 | \pi[X]\psi_2 \rangle.$$

This implies (6.10). Taking $X = I$ and using the unitality of π , we obtain that U is an isometric map of \mathcal{K} into $\mathcal{H} \otimes \mathcal{H}_E$.

Exercise 6.11. Show that the image of \mathcal{K} under the map U coincides with $\mathcal{H} \otimes \mathcal{H}_E$. □

The Stinespring representation (6.6) for a given CP map is not unique. Representations for which $\dim \mathcal{H}_E$ is minimal are called *minimal*. In this case, the number $\dim \mathcal{H}_E$ is called the *rank* of the map Φ .

Theorem 6.12. *Let*

$$\Phi^*[X] = \tilde{V}^*(X \otimes \tilde{I}_E)\tilde{V}, \quad X \in \mathfrak{B}(\mathcal{H}_B) \quad (6.13)$$

be another Stinespring representation, where \tilde{I}_E is the unit operator in $\tilde{\mathcal{H}}_E$. Then there exists a partial isometry W_E from \mathcal{H}_E to $\tilde{\mathcal{H}}_E$, such that

$$(I_B \otimes W_E)V = \tilde{V}; \quad (6.14)$$

if both representations are minimal, then W_E isometrically maps \mathcal{H}_E onto $\tilde{\mathcal{H}}_E$.

Proof. For the representation (6.6), consider the subspace

$$\mathcal{M} = \{(X \otimes I_E)V\psi : \psi \in \mathcal{H}_A, X \in \mathfrak{B}(\mathcal{H}_B)\} \subset \mathcal{K} = \mathcal{H}_B \otimes \mathcal{H}_E.$$

It is invariant under multiplication by operators of the form $Y \otimes I_E$. Hence, it has the form $\mathcal{M} = \mathcal{H}_B \otimes \mathcal{M}_E$, where $\mathcal{M}_E \subseteq \mathcal{H}_E$. For a minimal representation we should have $\mathcal{M}_E = \mathcal{H}_E$, because otherwise there would be a proper sub-representation.

Consider a similar subspace $\tilde{\mathcal{M}} = \mathcal{H}_B \otimes \tilde{\mathcal{M}}_E$ of the space $\tilde{\mathcal{K}} = \mathcal{H}_B \otimes \tilde{\mathcal{H}}_E$ for the second representation (6.13). Define the operator W from \mathcal{M} to $\tilde{\mathcal{M}}$ by the relation

$$W(X \otimes I_E)V\psi = (X \otimes \tilde{I}_E)\tilde{V}\psi. \quad (6.15)$$

In this case, W is isometric, since the norms of the vector $(X \otimes I_E)V\psi$ and of its image under W are both equal to $\langle \psi | \Phi[X^*X] | \psi \rangle$ by (6.6), (6.13).

From (6.15) we obtain, for all $Y \in \mathfrak{B}(\mathcal{H}_B)$,

$$W(YX \otimes I_E)V\psi = (Y \otimes \tilde{I}_E)W(X \otimes I_E)V\psi$$

and hence

$$W(Y \otimes I_E) = (Y \otimes \tilde{I}_E)W \quad (6.16)$$

on \mathcal{M} . Extend W to the whole of \mathcal{K} by setting it equal to zero on the orthogonal complement to \mathcal{M} . In this case (6.16) holds on \mathcal{K} . Therefore, $W = I_B \otimes W_E$, where W_E isometrically maps \mathcal{M}_E onto $\tilde{\mathcal{M}}_E$. Relation (6.15) implies (6.14). If the first representation is minimal, W_E is an isometry of \mathcal{H}_E into $\tilde{\mathcal{H}}_E$, and if both representations are minimal, W_E is unitary onto $\tilde{\mathcal{H}}_E$. \square

Corollary 6.13. *A map Φ^* is completely positive if and only if it can be represented in the form*

$$\Phi^*[X] = \sum_{k=1}^{\tilde{d}} V_k^* X V_k, \quad X \in \mathfrak{B}(\mathcal{H}_B), \quad (6.17)$$

where $V_k : \mathcal{H}_A \rightarrow \mathcal{H}_B$, or dually

$$\Phi[S] = \sum_{k=1}^{\tilde{d}} V_k S V_k^*, \quad S \in \mathfrak{T}(\mathcal{H}_A). \quad (6.18)$$

The map Φ^* is unital (Φ is trace preserving) if and only if

$$\sum_{k=1}^{\tilde{d}} V_k^* V_k = I. \quad (6.19)$$

This representation is often called the *Kraus representation*.

Proof. By writing $I_E = \sum_{k=1}^{\tilde{d}} |e_k^0\rangle\langle e_k^0|$, where $\{e_k^0\}$ is an orthonormal basis in \mathcal{H}_E and $\tilde{d} = \dim \mathcal{H}_E$, consider the operators V_k acting from \mathcal{H}_A to \mathcal{H}_B , defined by

$$\langle \phi | V_k | \psi \rangle = \langle \phi \otimes e_k^0 | V | \psi \rangle, \quad \phi \in \mathcal{H}_B, \quad \psi \in \mathcal{H}_A. \quad (6.20)$$

For the operators defined like in (6.20) we shall sometimes use the notation $V_k = \langle e_k^0 | V$. The representation (6.17) then follows from the formula (6.6). \square

Given a representation (6.17) or (6.18), we can always restore the Stinespring representation by letting $\mathcal{H}_E = \mathbb{C}^{\tilde{d}}$ and

$$V|\psi\rangle = \sum_{k=1}^{\tilde{d}} V_k|\psi\rangle \otimes |e_k^0\rangle.$$

Not surprisingly, the Kraus representation of a given completely positive map is not unique. Representations with a minimal number of components come from the minimal Stinespring representations and are also called *minimal*.

Exercise 6.14. Prove the following statement: the Kraus representation is minimal if and only if the operators $\{V_k\}$ are linearly independent. Hint: from (6.20)

$$\langle \phi | X \left(\sum_{k=1}^{\tilde{d}} c_k V_k \right) | \psi \rangle = \langle \phi | \otimes \left(\sum_{k=1}^{\tilde{d}} \bar{c}_k e_k^0 \right) | (X \otimes I_E) V | \psi \rangle,$$

for all $\phi \in \mathcal{H}_B$, $\psi \in \mathcal{H}_A$, $X \in \mathfrak{B}(\mathcal{H}_B)$. Hence, $\sum_{k=1}^{\tilde{d}} c_k V_k = 0$ is equivalent to $\sum_{k=1}^{\tilde{d}} \bar{c}_k e_k^0 \perp \mathcal{M}_E$.

Exercise 6.15. Basing yourself on Theorem 6.12, show that for any two Kraus representations of the same completely positive map, with the operators V_j, \tilde{V}_k , there exists a rectangular, partially isometric matrix $[u_{kj}]$, such that $\tilde{V}_k = \sum_j u_{kj} V_j$. If both representations are minimal, the matrix is unitary.

Exercise 6.16. Let $M = \{M_x\}$ be an observable. Show that if

$$M_x = V^* E_x V = \tilde{V}^* \tilde{E}_x \tilde{V} \quad (6.21)$$

are two Naimark's dilations of M in $\mathcal{K}, \tilde{\mathcal{K}}$, then there exists a partial isometry $W : \mathcal{K} \rightarrow \tilde{\mathcal{K}}$, such that

$$WV = \tilde{V}, \quad WE_x = \tilde{E}_x W.$$

If the dilations are minimal in the sense of the dimensionality of $\mathcal{K}, \tilde{\mathcal{K}}$, then W is isometry of \mathcal{K} onto $\tilde{\mathcal{K}}$.

The set of all CP maps that satisfy the normalization condition (6.19) is apparently convex. Extreme points of this set are given by

Proposition 6.17 (Choi [41]). *The unital CP map Φ^* is extreme if and only if it has a representation (6.17) where the system $\{V_j^* V_k\}$ is linearly independent.*

Proof. Let Φ^* be extreme and let (6.17) be its minimal representation. Let $\sum_{jk} y_{jk} V_j^* V_k = 0$ and denote $Y = \sum_{jk} y_{jk} |e_j^0\rangle\langle e_k^0|$. Without loss of generality, we can assume that $Y = Y^*$ and $\|Y\| \leq 1$. Define

$$\Phi_{\pm}^*[X] = V^*(X \otimes (I_B \pm Y))V = \sum_{jk} (\delta_{jk} \pm y_{jk}) V_j^* X V_k.$$

In this case, Φ_{\pm}^* are completely positive maps and $\Phi_{\pm}^*[I_B] = I_A$ by the assumption. Since Φ^* is extreme, one has $\Phi^* = \Phi_{\pm}^*$, implying $V^*(X \otimes Y)V = 0$ for all X, Y . It follows that

$$\langle \psi_2 | V^*(X_2^* \otimes I_E)(I_B \otimes Y)(X_1 \otimes I_E)V | \psi_1 \rangle \equiv 0.$$

Hence, the bilinear form of the operator $I_B \otimes Y$ vanishes on \mathcal{M} and hence $Y = 0$, by the minimality of the representation. Thus, $y_{jk} \equiv 0$, i.e. the system $\{V_j^* V_k\}$ is linearly independent.

Conversely, let the system $\{V_j^* V_k\}$ be linearly independent. Let $\Phi^* = \frac{1}{2}(\Phi_1^* + \Phi_2^*)$, where Φ_j^* are unital CP maps. In this case, the Kraus operators for Φ_1^*, Φ_2^* are also Kraus operators for some representation of Φ^* , and hence are linearly expressible through $\{V_k\}$, by Exercise 6.15, so that $\Phi_1^*[X] = \sum_{jk} z_{jk} V_j^* X V_k$. Since $\Phi_1^*[I_B] = I_A$, we have

$$\sum_{jk} z_{jk} V_j^* V_k = \sum_{jk} \delta_{jk} V_j^* V_k.$$

Hence, $z_{jk} = \delta_{jk}$ by linear independence. Therefore, $\Phi_1^* = \Phi^*$ and thus Φ^* is extreme. \square

By Exercise 1.5, the dimensionality of $\mathfrak{B}(\mathcal{H}_A)$ is d_A^2 . Hence, the rank \tilde{d} of an extreme map should satisfy $\tilde{d}^2 \leq d_A^2$, i.e. $\tilde{d} \leq d_A$ as compared to $\tilde{d} \leq d_A d_B$ (see (6.9)) for an arbitrary map.

6.3 Definition of the channel

Let us provide a physical interpretation of the property of complete positivity, via the picture of the (irreversible) evolution of an open system interacting with the environment. Let \mathcal{H} be the Hilbert space of the system, \mathcal{H}_E the Hilbert space of its environment, and S_E the initial state of the environment. Assume that the system interacts with the environment via a unitary operator U . The evolution of the system is then given by the formula

$$\Phi[S] = \text{Tr}_E U (S \otimes S_E) U^*. \quad (6.22)$$

Theorem 6.18. *Every linear trace that preserves a completely positive map, with $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}$, can be extended to the evolution of an open system interacting with an environment so that relation (6.22) holds.*

Proof. Consider the space \mathcal{H}_E of \tilde{d} -dimensional vectors, where \tilde{d} is the number of components in the Kraus representation of Φ . In this case, $\mathcal{H} \otimes \mathcal{H}_E$ can be considered as the direct sum of \tilde{d} copies of \mathcal{H} , consisting of column vectors $[\psi_1, \dots, \psi_{\tilde{d}}]^\top$, $\psi_j \in \mathcal{H}$, while operators in this space are the block matrices $[X_{jk}]_{j,k=1,\dots,\tilde{d}}$, $X_{jk} \in \mathfrak{B}(\mathcal{H})$. Consider the vector $|\psi_E\rangle = [1, 0, \dots, 0]^\top$. In this case,

$$S \otimes |\psi_E\rangle\langle\psi_E| = \begin{bmatrix} S & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \end{bmatrix}.$$

Introduce the operator

$$U = \begin{bmatrix} V_1 & \dots & \dots \\ \dots & \dots & \dots \\ V_{\tilde{d}} & \dots & \dots \end{bmatrix}, \quad (6.23)$$

where the first column consists of $V_1, \dots, V_{\tilde{d}}$ and the rest of the columns are chosen to complement it to a unitary operator, which is possible due to (6.19). Now,

$$U (S \otimes |\psi_E\rangle\langle\psi_E|) U^* = [V_j S V_k^*]_{j,k=1,\dots,\tilde{d}}.$$

The partial trace in $\mathcal{H} \otimes \mathcal{H}_E$, with respect to \mathcal{H}_E , is just the sum of the diagonal elements of this matrix, which coincides with the Kraus representation for the map Φ . \square

The preceding discussion motivates the following definition. We denote by A the input system of the channel and by B the output system.

Definition 6.19. A *channel* in the Schrödinger picture is a linear, completely positive trace-preserving map $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$. Dually, a channel in the Heisenberg picture is a linear, completely positive unital map $\Phi^* : \mathfrak{B}(\mathcal{H}_B) \rightarrow \mathfrak{B}(\mathcal{H}_A)$.

The channel Φ is called *bistochastic* if it maps the chaotic state in \mathcal{H}_A into the chaotic state in \mathcal{H}_B , $\Phi[I_A/d_A] = I_B/d_B$. If the input and output systems have equal dimensionalities ($d_A = d_B$), the bistochastic channel Φ is unital. In this case, Φ^* is also trace-preserving, so that both Φ and Φ^* are channels in both the Schrödinger and Heisenberg pictures.

Let $S = S_A$ be an input state of a channel Φ . Denote by $\mathcal{L} = \text{supp } S$ the support of the density operator S . Consider a purification $S_{AR} = |\psi_{AR}\rangle\langle\psi_{AR}|$ of the state S_A , where R labels the purifying system and

$$|\psi_{AR}\rangle = \sum_j \sqrt{\lambda_j} |e_j\rangle \otimes |h_j\rangle; \quad \lambda_j > 0, e_j \in \mathcal{L},$$

cf. (3.9). Let Id_R be the identity map in $\mathfrak{T}(\mathcal{H}_R)$. In this case, the tensor product $\Phi \otimes \text{Id}_R$ describes a channel from AR to BR , transforming A into B and leaving R intact. Therefore, R is called the *reference* system. The output state of this channel is

$$(\Phi \otimes \text{Id}_R)[S_{AR}] = S_{BR}. \quad (6.24)$$

Since the state S_{AR} is, in general, entangled, this operation is sometimes called “entanglement transmission”. Notice that by the trace preservation property of Φ , the state S_{BR} has the same partial state S_R as the input state S_{AR} . Since S_R is a copy of the state S_A , the state S_{BR} can be considered a quantum substitute of the joint distribution of the input A and the output B of the channel Φ .

Proposition 6.20. *The state S_{BR} uniquely determines the restriction $\Phi_{\mathcal{L}}$ of the channel Φ to the states S with $\text{supp } S \subset \mathcal{L}$. In particular, if S_A is nondegenerate, S_{BR} uniquely determines the channel Φ .*

Proof. One has

$$(\Phi \otimes \text{Id}_R)[S_{AR}] = \sum_{j,k} \sqrt{\lambda_j} \sqrt{\lambda_k} \Phi[|e_j\rangle\langle e_k|] \otimes |h_j\rangle\langle h_k|. \quad (6.25)$$

The collection of operators $\Phi[|e_j\rangle\langle e_k|]$ uniquely determines $\Phi_{\mathcal{L}}$. Hence, the statement follows. \square

In particular, taking the maximally entangled state

$$S_{AR} = \frac{1}{d} \sum_{j,k} |e_j\rangle\langle e_k| \otimes |h_j\rangle\langle h_k|,$$

we obtain from (6.25), for arbitrary state S ,

$$\Phi[S] = d \text{Tr}_R S_{BR} (I_B \otimes S_R^\top), \quad (6.26)$$

where $S_R^\top = \sum_{j,k} \langle e_k | S | e_j \rangle |h_j\rangle\langle h_k|$ and the state S_{BR} satisfies the characteristic condition $\text{Tr}_B S_{BR} = \bar{S}_R = I_R/d$. This one-to-one correspondence between channels and states is called the *Choi–Jamiołkowski correspondence*.

6.4 Entanglement-breaking and PPT channels

In Quantum Information Theory one often has to deal with both quantum and classical information. A usual device is to embed the classical system into a quantum system, as in Section 2.1.1, by representing classical states and observables on the phase space Ω as diagonal operators in the artificial Hilbert space \mathcal{H} spanned by the orthonormal basis $\{|\omega\rangle; \omega \in \Omega\}$.

- i. Any classical-quantum (c-q) channel $x \rightarrow S_x$ describing an encoding of the classical input x into the quantum state S_x , as considered in Chapter 5, can then be extended to the quantum channel,

$$\Phi[S] = \sum_x \langle e_x | S | e_x \rangle S_x, \quad S \in \mathfrak{S}(\mathcal{H}_A),$$

where $\{e_x\}$ is an orthonormal basis in \mathcal{H}_A .

- ii. The quantum-classical (q-c) channel corresponding to the measurement of observable $M = \{M_y\}$ in \mathcal{H}_A (see Section 6.5 below), which produces the probability distribution from a quantum state, gives rise to the quantum channel

$$\Phi[S] = \sum_y |e_y\rangle\langle e_y| \text{Tr } S M_y, \quad S \in \mathfrak{S}(\mathcal{H}_A),$$

where $\{e_y\}$ is an orthonormal basis in \mathcal{H}_B . The dual channel acts as

$$\Phi^*[X] = \sum_y \langle e_y | X | e_y \rangle M_y. \quad (6.27)$$

The Stinespring representation (6.6) in this case provides Naimark's dilation for the observable M . By taking $X = |e_y\rangle\langle e_y|$, we obtain

$$M_y = V^* (|e_y\rangle\langle e_y| \otimes I_E) V,$$

where $E = \{|e_y\rangle\langle e_y| \otimes I_E\}$ is the sharp observable.

- iii. Taking a composition of q-c and c-q channels with the same orthonormal basis $\{e_j\}$, we obtain the q-c-q channel

$$\Phi[S] = \sum_j S_j \text{Tr } S M_j, \quad (6.28)$$

of which i. and ii. are particular cases. Such channels have a classical system between the input and the output (represented by symbol j).

Definition 6.21. Channel Φ from A to B is *entanglement-breaking* if for an arbitrary input state S_{AR} of the channel $\Phi \otimes \text{Id}$, its output state S_{BR} is separable (Definition 3.13), i.e. a mixture of product states

$$S_{BR} = \sum_j p_j S_B^j \otimes S_R^j, \quad (6.29)$$

where $\{p_j\}$ is a probability distribution.

The following characterization is due to M. Horodecki, Shor and Ruskai [121].

Proposition 6.22. *The following properties are equivalent:*

- i. channel Φ is *entanglement-breaking*
- ii. the state S_{BR} in the Choi–Jamiołkowski representation (6.26) is *separable*
- iii. Φ is *q-c-q channel* (6.28)

Proof.

i. \Rightarrow ii. is trivial.

ii. \Rightarrow iii. Taking for S_{AR} the maximally entangled state, and using relations (6.29) and (6.26), we obtain the representation (6.28), in which $S_j = S_B^j$, $M_j = dp_j(S_R^j)^\top$.

- iii. \Rightarrow i. For a channel (6.28) and any input state S_{AR} of the channel $\Phi \otimes \text{Id}_R$, relation (6.29) holds, with $S_B^j = S_j$, $p_j S_R^j = \text{Tr}_A S_{AR}(M_j \otimes I_R)$. \square

Example 6.23. An important extreme case of a c-q channel is the *completely depolarizing* channel, describing the irreversible evolution to some final state S_f :

$$\Phi[S] = S_f \cdot \text{Tr } S, \quad S \in \mathfrak{T}(\mathcal{H}). \quad (6.30)$$

Exercise 6.24. Prove complete positivity and find a Kraus representation in the examples above.

Definition 6.25. Channel Φ from A to B is *PPT* if, for an arbitrary input state S_{AR} of the channel $\Phi \otimes \text{Id}$, its output state S_{BR} has a positive partial transpose in the space \mathcal{H}_R (see the end of Section 3.1.3).

Apparently, every entanglement-breaking channel is PPT (but not vice-versa, see Horodeckis' paper [118]; a PPT channel which is not entanglement-breaking is called *entanglement-binding*). If the channel Φ is given by the Kraus decomposition (6.18), the channel Φ^\top is defined by the relation

$$\Phi^\top[S] = \sum_k V_k^{*\top} S V_k^\top$$

assuming some bases are fixed in the input and output spaces of the channel. Taking into account that $S_{AR}^\top = S_{AR}$, we obtain from (6.24), for the transposition in the basis $\{e_j \otimes h_k\}$,

$$(\Phi^\top \otimes \text{Id}_R)[S_{AR}] = S_{BR}^\top.$$

It follows that the channels Φ and Φ^\top are simultaneously PPT or not PT.

Exercise 6.26. Prove the relation

$$\Phi \circ T_A = T_B \circ \Phi^\top, \quad (6.31)$$

where T_A, T_B denote transpositions in the corresponding spaces.

Proposition 6.27. *The following conditions are equivalent:*

- i. *The channel Φ is PPT.*
- ii. *The state S_{BR} in the Choi–Jamiołkowski representation (6.26) has a positive partial transpose.*
- iii. *$\Phi \circ T_A$ or $T_B \circ \Phi$ is a channel.*

Proof. Again i. \Rightarrow ii. is trivial.

ii. \Rightarrow iii. Taking for S_{AR} the maximally entangled state and using the relation (6.26), we obtain

$$(\Phi \circ T_A)[S] = d \operatorname{Tr}_R S_{BR} (I_B \otimes S_R) = d \operatorname{Tr}_R S_{BR}^{\top R} (I_B \otimes S_R^{\top}),$$

which is a channel, because $S_{BR}^{\top R} \geq 0$. By relation (6.31) and the remark preceding it, this is equivalent to $T_B \circ \Phi$ being a channel.

iii. \Rightarrow i. Let $T_B \circ \Phi$ be a channel. In this case, for any input state S_{AR} of the channel $\Phi \otimes \operatorname{Id}_R$ we have

$$(\Phi \otimes \operatorname{Id}_R[S_{AR}])^{\top R} = (T_B \circ \Phi \otimes \operatorname{Id}_R)[S_{AR}]^{\top} \geq 0. \quad \square$$

6.5 Quantum measurement processes

An important example of irreversible evolution is the state change of the system due to a quantum measurement. A *complete ideal* quantum measurement is associated with an orthonormal basis $|e_x\rangle$, indexed by the measurement outcomes x . It is postulated that the initial state S of the system after such a measurement is transformed into one of the states $|e_x\rangle\langle e_x|$ with probability $\langle e_x|S|e_x\rangle$. Thus, the post-measurement statistical ensemble splits into subensembles corresponding to different measurement outcomes x , and, as a whole, is described by the density operator

$$S' = \sum_x |e_x\rangle\langle e_x|S|e_x\rangle\langle e_x|. \quad (6.32)$$

Notice that the map $S \rightarrow S'$ is a particular case of a q-c channel corresponding to the measurement of an observable $M = \{|e_x\rangle\langle e_x|\}$.

Under an *incomplete* ideal measurement some outcomes are joined together (coarse-grained), forming the orthogonal resolution of the identity $E = \{E_x\}$,

$$E_x E_{x'} = \delta_{xx'} E_x, \quad \sum_x E_x = I,$$

in other words, a sharp observable in the space \mathcal{H} . According to the *von Neumann–Lüders projection postulate*, the ideal measurement of the sharp observable E gives an outcome x with the probability

$$p_x = \operatorname{Tr} S E_x = \operatorname{Tr} E_x S E_x,$$

and the *posterior state*, i.e. the post-measurement state of the subensemble in which the outcome x has occurred is

$$S_x = \frac{E_x S E_x}{\operatorname{Tr} E_x S E_x}, \quad \text{if } p_x > 0.$$

We then have the quantum analog of the *Bayes formula* for the state of the whole ensemble

$$S' = \sum_x p_x S_x = \sum_x E_x S E_x = \mathcal{E}[S]. \quad (6.33)$$

Note that, as distinct from the classical Bayes formula, the quantum state (6.33) of the whole ensemble after the ideal measurement can be different from the initial state S . Thus, even an ideal quantum measurement does not reduce to a simple reading of the outcome and involves an inevitable interaction, which changes the state of the system. This is a fundamental difference between quantum observables and classical random variables, the observation of which does not change the statistical ensemble and reduces to mere selection of its representatives according to the values of a random variable.

In the Heisenberg picture, the transformation of observables due to an ideal measurement is described by the completely positive map

$$\mathcal{E}^*[X] = \sum_x E_x X E_x = \mathcal{E}[X], \quad (6.34)$$

which can be interpreted as a conditional expectation onto the algebra of operators commuting with all of E_x , characterized by the properties

- i. $\mathcal{E}^2 = \mathcal{E}$
- ii. $\mathcal{E}[X \mathcal{E}[Y]] = \mathcal{E}[X] \mathcal{E}[Y]$, $X, Y \in \mathfrak{B}(\mathcal{H})$.

An ideal quantum measurement satisfies the *repeatability hypothesis*: the outcome of the repeated measurement is equal, with probability one, to the outcome of the first measurement (provided no evolution occurred between the two measurements). Most of the real measurement procedures do not fulfill this requirement, and we pass to a mathematical description of non-ideal measurements.

Consider the process of *indirect* measurement, when the system \mathcal{H} in the state S first interacts with a probe system \mathcal{H}_E in the initial state S_E , and next an ideal measurement of a sharp observable $\{E_x^0\}$ is made over the probe system. In this case, according to the above, the probability of an outcome x equals

$$p_x = \text{Tr } U(S \otimes S_E) U^*(I \otimes E_x^0),$$

and the posterior state is described by the density operator S_x , such that

$$p_x S_x = \text{Tr}_E U(S \otimes S_E) U^*(I \otimes E_x^0) \equiv \Phi_x[S]. \quad (6.35)$$

Thus, the state of the system after the measurement can be written as

$$\Phi[S] = \sum_x p_x S_x = \sum_x \Phi_x[S].$$

The maps Φ_x are completely positive, and their sum Φ is trace preserving. Any such family $\{\Phi_x\}$ is called an *instrument*. The instruments describe the statistics and posterior states of non-ideal measurements, which in general need not satisfy the repeatability hypothesis.

Note that

$$p_x = \text{Tr } S M_x, \quad \text{where} \quad M_x = \Phi_x^*[I]$$

is the resolution of the identity describing the observable associated with the instrument and the indirect measurement. Conversely, given an observable $M = \{M_x\}$, one can construct a (non-unique) instrument and indirect measurement with which it is associated. Consider a decomposition $M_x = V_x^* V_x$ and the corresponding unitary operator (6.23) in the space $\mathcal{H} \otimes \mathcal{H}_E$. In this case, the family of completely positive maps

$$\Phi_x[S] = \text{Tr}_E U(S \otimes |\psi_E\rangle\langle\psi_E|) U^*(I \otimes |e_x\rangle\langle e_x|), \quad (6.36)$$

where $|e_x\rangle$ is the column vector with 1 in the x -th place and zeroes otherwise, so that $|\psi_E\rangle = |e_1\rangle$, forms an instrument. Further,

$$M_x = \text{Tr}_E (I \otimes |\psi_E\rangle\langle\psi_E|) U^*(I \otimes |e_x\rangle\langle e_x|) U = \Phi_x^*[I]. \quad (6.37)$$

Thus, the process of indirect measurement of the observable M is realized by the probe system \mathcal{H}_E in the initial state $|\psi_E\rangle\langle\psi_E|$, the unitary operator U in the space $\mathcal{H} \otimes \mathcal{H}_E$, and the sharp observable $\{|e_x\rangle\langle e_x|\}$ in \mathcal{H}_E .

Exercise 6.28. For a complete ideal measurement with $M_x = |e_x\rangle\langle e_x|$, set $V_x = M_x$ and construct the unitary operator U by use of formula (6.23) (in this case $\mathcal{H}_E = \mathcal{H}$). Show that

$$U(|e_x\rangle \otimes |e_1\rangle) = |e_x\rangle \otimes |e_x\rangle.$$

For an arbitrary vector $\psi \in \mathcal{H}$

$$U(|\psi\rangle \otimes |e_1\rangle) = \sum_x \langle e_x | \psi \rangle (|e_x\rangle \otimes |e_x\rangle),$$

thus, an interaction U produces entanglement between the main and the probe systems, which allows us to reduce the ideal measurement in the main system to the measurement of the sharp observable $\{|e_x\rangle\langle e_x|\}$ in the probe system.

6.6 Complementary channels

Given three quantum systems A, B, C with the spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C$ and a linear operator $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_C$, the relations

$$\Phi_B[S] = \text{Tr}_C V S V^*, \quad \Phi_C[S] = \text{Tr}_B V S V^*; \quad S \in \mathfrak{T}(\mathcal{H}_A) \quad (6.38)$$

define two CP maps $\Phi_B : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$, $\Phi_C : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_C)$, which will be called mutually *complementary*. If V is an isometry, both maps are trace preserving i.e. channels.

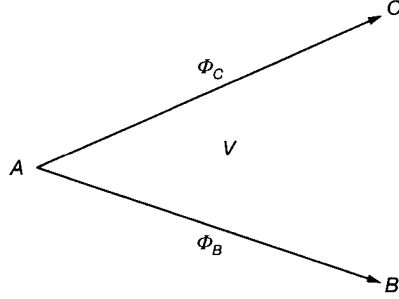


Figure 6.1. Complementary channel.

The Stinespring dilation theorem implies that for a given CP map (channel) a complementary always exists. Moreover, the complementary map is unique in the following sense: for a given CP map Φ_B , any two maps $\Phi_C, \Phi_{C'}$ complementary to Φ_B are isometrically equivalent in the sense that there is a partial isometry $W : \mathcal{H}_C \rightarrow \mathcal{H}_{C'}$ such that

$$\Phi_{C'}(S) = W \Phi_C(S) W^*, \quad \Phi_C(S) = W^* \Phi_{C'}(S) W, \quad (6.39)$$

for all S . This is a consequence of Theorem 6.12. If the dimensionalities of $\mathcal{H}_C, \mathcal{H}_{C'}$ are minimal, W is an isometry from \mathcal{H}_C onto $\mathcal{H}_{C'}$. In this case, the complementary map is called *minimal*.

If $A = B$ is an open quantum system that interacts with the environment $C = E$ in an arbitrary initial state S_E , and Φ_B is the corresponding channel

$$\Phi_B[S] = \text{Tr}_E U(S \otimes S_E) U^*, \quad (6.40)$$

describing the state change of A , the final state of the environment is the output of the channel

$$\Phi_E[S] = \text{Tr}_B U(S \otimes S_E) U^*. \quad (6.41)$$

We will call the channels Φ_B, Φ_E mutually *weakly complementary*. If the state of the environment is pure, $S_E = |\psi_E\rangle\langle\psi_E|$, then introducing the isometry $V = U|\psi_E\rangle : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$, which acts as follows

$$V|\psi\rangle = U(|\psi\rangle \otimes |\psi_E\rangle), \quad \psi \in \mathcal{H}_A,$$

we see that Φ_E is the complementary of Φ_B . If S_E is not pure, then by taking its purification $S_{E'}$ in the space $\mathcal{H}_{E'} = \mathcal{H}_E \otimes \mathcal{H}_R$ and letting $U' = U \otimes I_R$, we obtain a representation of the type (6.41), with E replaced by E' , where $S_{E'}$ is a pure state.

Hence, $\Phi_{E'}$ is the complementary of Φ_B and $\Phi_E[S] = \text{Tr}_R \Phi_{E'}[S]$. In contrast to the complementary map, the weak complementary is not unique and depends on the representation (6.40).

To simplify the formulas we shall also use the notation $\tilde{\Phi}$ for the map that is complementary to Φ .

Assume that a CP map $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ is given by the Kraus representation

$$\Phi[S] = \sum_{k=1}^{\tilde{d}} V_k S V_k^*.$$

In this case, a complementary map $\tilde{\Phi} : \mathfrak{T}(\mathcal{H}) \rightarrow \mathfrak{M}_{\tilde{d}}$ is given by

$$\tilde{\Phi}[S] = [\text{Tr } V_k S V_l^*]_{k,l=1,\tilde{d}} = \sum_{k,l=1}^{\tilde{d}} (\text{Tr } S V_l^* V_k) |e_k\rangle\langle e_l|, \quad (6.42)$$

where $\{e_k\}$ is the canonical basis for the coordinate space $\mathbb{C}^{\tilde{d}}$, which plays the role of \mathcal{H}_C , so that $\tilde{d} = d_C$. This statement follows from the fact that

$$V = \sum_{k=1}^{\tilde{d}} \oplus V_k$$

is a map from \mathcal{H}_A to $\sum_{k=1}^{\tilde{d}} \oplus \mathcal{H}_B \simeq \mathcal{H}_B \otimes \mathcal{H}_C$, for which $\Phi, \tilde{\Phi}$ are given by the partial traces (6.38), see Exercise 3.6.

Exercise 6.29. Check by direct computation that, applying the same procedure to $\tilde{\Phi}$, one obtains the map $\tilde{\tilde{\Phi}}$, which is isometrically equivalent in the sense of (6.39) to Φ .

Exercise 6.30. Prove the following statement: channel Φ is extreme if and only if there exists a complementary channel $\tilde{\Phi}$ such that $\text{Ker } \tilde{\Phi}^* = 0$. Hint: prove the relation $\tilde{\Phi}^*[Y] = \sum_{jk} y_{jk} V_j^* V_k$, where $y_{jk} = \langle e_j^0 | Y | e_k^0 \rangle$ and use Proposition 6.17.

Example 6.31. Consider the channel

$$\Phi(S) = S \otimes S_C, \quad (6.43)$$

where S_C is a fixed state in the space \mathcal{H}_C . By writing the spectral decomposition

$$S_C = \sum_{k=1}^{\tilde{d}} \lambda_k |e_k\rangle\langle e_k|,$$

one arrives at the Kraus representation with the operators $V_k = \sqrt{\lambda_k}(I \otimes |e_k\rangle) : \mathcal{H}_A \rightarrow \mathcal{H}_A \otimes \mathcal{H}_C$. Hence, one obtains from (6.42)

$$\tilde{\Phi}(S) = [\lambda_k \delta_{kl} \text{Tr} S]_{k,l=1,\bar{d}} \simeq S_C \text{Tr} S,$$

which is a completely depolarizing channel. If S_C is pure, (6.43) is essentially the same as the ideal (identity) channel Id and $\tilde{\Phi}$ is its complementary. Whatever the state S_C , the ideal channel Id is weakly complementary for this channel.

This example illustrates the fact that perfect transmission of quantum information from input A to output B is equivalent to no information going from A to the environment C , and vice versa. Later in this book, we shall see that approximate version of this complementarity plays a basic role in quantum capacity theorems: the closer channel Φ is to an ideal one, the closer its complement $\tilde{\Phi}$ is to the completely depolarizing one.

The completely depolarizing channel is an instance of entanglement-breaking (q-c-q) channel (see Section 6.4); the construction of the complementary channel can be generalized to the whole of this class. For this, we need the following characterization of entanglement-breaking channels.

Proposition 6.32. *Channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ is entanglement-breaking if and only if it has a Kraus representation with rank one operators V_k :*

$$\Phi[S] = \sum_{k=1}^{\bar{d}} |\varphi_k\rangle\langle\varphi_k| S |\psi_k\rangle\langle\psi_k|, \quad (6.44)$$

where

$$\sum_{k=1}^{\bar{d}} |\psi_k\rangle\langle\varphi_k| \langle\varphi_k| \langle\psi_k| = I.$$

Proof. Let the channel Φ admit the representation (6.44). In this case, making a renormalization, we can always assume that $\langle\varphi_k|\varphi_k\rangle = 1$, in which case $\{|\psi_k\rangle\}$ is an overcomplete system. Denoting $S_k = |\varphi_k\rangle\langle\varphi_k|$ and $M_k = |\psi_k\rangle\langle\psi_k|$, we rewrite (6.44) in the form (6.28). Hence, Φ is entanglement-breaking.

Conversely, by making the spectral decompositions of the operators S_j, M_j in (6.28), we obtain a Kraus representation of the form (6.44). \square

The complementary channel $\tilde{\Phi} : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathbb{C}^{\bar{d}})$ has the form

$$\tilde{\Phi}[S] = [c_{kl} \langle\psi_k| S |\psi_l\rangle]_{k,l=1,\bar{d}} = \sum_{k,l=1}^{\bar{d}} c_{kl} |e_k\rangle\langle\psi_k| S |\psi_l\rangle\langle e_l|, \quad (6.45)$$

where $c_{kl} = \langle \varphi_l | \varphi_k \rangle$. In the special case where the system $\{\psi_k\}_{k=1, \bar{d}}$ is an orthonormal basis in \mathcal{H}_A , (6.45) is a channel called *dephasing*. The reason for this name is that such channels describe the suppression of the off-diagonal elements of the density matrix, due to the fact that $|c_{kl}| \leq 1$. In particular, the *completely dephasing* channel corresponds to $c_{kl} = \delta_{kl}$. From (6.44) we see that the dephasing channels are complementary to c-q channels. Note that using the Schur (element-wise) product of matrices $*$, formula (6.45) can in this case be written in the compact form

$$\tilde{\Phi}[S] = C * S,$$

where $C = [c_{kl}]$ and S is the density matrix in the orthonormal basis $\{\psi_k\}$.

For q-c channels, $\{\varphi_k\}_{k=1, \bar{d}}$ is an orthonormal basis in \mathcal{H} , so that $c_{kl} = \delta_{kl}$, and the complementary channel is again a q-c channel.

Noting that an arbitrary positive definite matrix $[c_{kl}]$ can be represented as $c_{kl} = \sum_j \tilde{v}_{lj} v_{kj}$, and denoting

$$\tilde{V}_j = \sum_{k=1}^{\bar{d}} v_{kj} |e_k\rangle \langle \psi_k|, \quad (6.46)$$

we have the Kraus representation

$$\tilde{\Phi}[S] = \sum_{j=1}^{d_B} \tilde{V}_j S \tilde{V}_j^* \quad (6.47)$$

for the complementary map. For the dephasing channel, we can take $|e_k\rangle = |\psi_k\rangle$. Hence, one sees from (6.46) that the dephasing maps are characterized by the property of having a Kraus representation with simultaneously diagonalizable (i.e. commuting normal) operators \tilde{V}_j .

Definition 6.33. Let $\{p_j\}$ be a finite probability distribution and $\Phi_j : \mathfrak{T}(\mathcal{H}) \rightarrow \mathfrak{T}(\mathcal{H}'_j)$ a collection of channels. Channel $\Phi : \mathfrak{T}(\mathcal{H}) \rightarrow \mathfrak{T}(\sum_j \oplus \mathcal{H}'_j)$ will be called *orthogonal convex sum* of the channels Φ_j ,

$$\Phi = \sum_j \oplus p_j \Phi_j,$$

if $\Phi[S] = \sum_j \oplus p_j \Phi_j[S]$ for all $S \in \mathfrak{S}(\mathcal{H})$.

Example 6.34. *Quantum erasure channel.* Let $p \in [0, 1]$. Now, the channel $\Phi : \mathfrak{T}(\mathcal{H}) \rightarrow \mathfrak{T}(\mathcal{H} \oplus \mathbb{C})$, defined by the relation

$$\Phi(S) = \begin{bmatrix} (1-p)S & 0 \\ 0 & p \text{Tr } S \end{bmatrix}$$

transmits the input state S with probability $1 - p$ and “erases” it with probability p sending the erasure signal. This channel is the orthogonal convex sum of the ideal channel and the completely depolarizing channel (6.30), with a pure final state S_f .

Exercise 6.35. Show that the complementary to the orthogonal convex sum of the channels is the orthogonal convex sum of the complementary channels for the components:

$$\widetilde{\sum_j \oplus p_j \Phi_j} = \sum_j \oplus p_j \tilde{\Phi}_j.$$

Hint: write the Kraus representation for the orthogonal convex sum, basing yourself on those for the summands, and then use the relation (6.42).

In particular, by using Example 6.31 (with $\tilde{d} = \dim \mathcal{H}_C = 1$), find that the complementary to the erasure channel Φ_p is unitarily equivalent to the erasure channel Φ_{1-p} .

6.7 Covariant channels

Recall that a projective unitary representation of the group G is a family of unitary operators $\{V_g; g \in G\}$ in a Hilbert space, \mathcal{H} satisfying

$$V_g V_h = \lambda(g, h) V_{gh},$$

where $\lambda(g, h)$ is a complex multiplier (of modulus one). If $\lambda(g, h) \equiv 1$, we call it a unitary representation (see e.g. [153] for more detail).

Let G be a group and $g \rightarrow V_g^{(j)}$; $g \in G$ be two (projective) unitary representations of G in \mathcal{H}_j ; $j = A, B$. The channel $\Phi: \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ is *covariant* if

$$\Phi[V_g^A S (V_g^A)^*] = V_g^B \Phi[S] (V_g^B)^* \quad (6.48)$$

for all $g \in G$ and S . If the representation V_g^B is irreducible, the channel Φ is bistochastic. Indeed, $V_g^B \Phi[I_A] = \Phi[I_A] V_g^B$ (for all g), and irreducibility of V_g^B implies that $\Phi[I_A]$ is proportional to I_B , with the coefficient obtained by taking the traces.

Definition 6.36. The *depolarizing channel* in \mathcal{H} is defined by the relation

$$\Phi[S] = (1 - p)S + p \frac{I}{d} \text{Tr } S, \quad 0 \leq p \leq \frac{d^2}{d^2 - 1}. \quad (6.49)$$

If $p \leq 1$, this describes a mixture of the ideal channel Id and the completely depolarizing channel which maps an arbitrary state to the chaotic state $\tilde{S} = \frac{I}{d}$. For the whole range $0 \leq p \leq \frac{d^2}{d^2 - 1}$ complete positivity can be demonstrated by the explicit Kraus representation (see (6.54) below).

Exercise 6.37. Show that the depolarizing channel can be characterized by the property of unitary covariance: $\Phi[USU^*] = U\Phi[S]U^*$ for an arbitrary unitary operator U in \mathcal{H} .

The depolarizing channel satisfies $\Phi^* = \Phi$ and is thus unital and bistochastic.

Exercise 6.38. Show that the channel complementary to a covariant channel is itself covariant:

$$\Phi[V_g^A S (V_g^A)^*] = V_g^E \Phi[S] (V_g^E)^*,$$

where $g \rightarrow V_g^E$ is a (projective) unitary representation of the group G in \mathcal{H}_E . Hint: use the fact that an arbitrary covariant channel has a Kraus representation (6.18), in which operators V_j satisfy the relations

$$V_g^B V_j (V_g^A)^* = \sum_k d_{jk}(g) V_k,$$

where $g \rightarrow D(g) = [d_{jk}(g)]$ is a matrix unitary representation of the group G . It follows that the complementary map (6.42) is covariant and the role of the second representation V_g^E is played by $D(g)$.

For future use we introduce here an important general construction, namely a discrete version of the Weyl operators and canonical commutation relations (CCR) in a finite-dimensional Hilbert space. Fix an orthonormal basis $\{e_k; k = 1, \dots, d\}$ in \mathcal{H} . The *discrete Weyl operators* are the unitary operators in \mathcal{H} defined as

$$W_{\alpha\beta} = U^\alpha V^\beta; \quad \alpha, \beta = 0, \dots, d-1, \quad (6.50)$$

where

$$V|e_k\rangle = \exp\left(\frac{2\pi i k}{d}\right)|e_k\rangle; \quad U|e_k\rangle = |e_{k+1(\bmod d)}\rangle; \quad k = 0, \dots, d-1. \quad (6.51)$$

Notice that $W_{00} = I$. These operators satisfy the discrete analog of the Weyl–Segal canonical commutation relations (CCR) for bosonic systems, which will appear in Chapter 12:

$$\begin{aligned} W_{\alpha\beta} W_{\alpha'\beta'} &= \exp \frac{2\pi i \beta \alpha'}{d} W_{\alpha+\alpha', \beta+\beta'} \\ &= \exp \frac{2\pi i (\beta \alpha' - \beta' \alpha)}{d} W_{\alpha'\beta'} W_{\alpha\beta}. \end{aligned} \quad (6.52)$$

The first equality expresses the fact that $(\alpha, \beta) \rightarrow W_{\alpha\beta}$ is a projective representation of the additive cyclic group $\mathbb{Z}_d \times \mathbb{Z}_d$.

The representation is irreducible: any operator A that commutes with all $W_{\alpha,\beta}$ is a multiple of the unit operator. Indeed, $[A, V] = 0$ implies, by (6.51), that

$A|e_k\rangle = a_k|e_k\rangle$ and $[A, U] = 0$ implies $a_k \equiv a$. We then have an important identity: for all states S

$$\sum_{\alpha, \beta=0}^{d-1} W_{\alpha\beta} S W_{\alpha\beta}^* = d(\text{Tr } S)I, \quad (6.53)$$

since it follows from (6.52) that the left side commutes with all the Weyl operators. The constant on the right is obtained by comparing traces. In fact, the identity (6.53) is a very special case of a general identity, following from orthogonality relations for an arbitrary irreducible representation (see e.g. Section IV.2 in [107]).

Exercise 6.39. Prove the Kraus representation for the depolarizing channel (6.49):

$$\Phi[S] = \left(1 - p \frac{d^2 - 1}{d^2}\right) S + \frac{p}{d^2} \sum_{\alpha+\beta>0} W_{\alpha\beta} S W_{\alpha\beta}^*, \quad (6.54)$$

where all the coefficients are nonnegative, provided $0 \leq p \leq \frac{d^2}{d^2-1}$. Hint: substitute the expression for $\text{Tr } S$ from (6.53) into (6.49).

Proposition 6.40. *The depolarizing channel is entanglement-breaking for $p \geq \frac{d}{d+1}$.*

Proof. To avoid notational confusion with the differential, in this proof we denote the dimensionality as $d_A = \dim \mathcal{H}$. Let Θ be the unit sphere in \mathcal{H} and let $\nu(d\theta)$ be the uniform distribution on Θ . Thus, $|\theta\rangle \in \Theta$ is a unit vector in \mathcal{H} . Now, as we show below,

$$d_A \int_{\Theta} |\theta\rangle\langle\theta| \nu(d\theta) = I, \quad (6.55)$$

so that we have a *continuous overcomplete system* in \mathcal{H} . Then as we shall see in Section 11.7, the relation $M(d\theta) = d_A |\theta\rangle\langle\theta| \nu(d\theta)$ defines an observable in \mathcal{H} , with values in Θ , in the sense of Definition 11.29. Its probability distribution in the state S is given by

$$\mu_S^M(d\theta) = d_A \langle\theta|S|\theta\rangle \nu(d\theta).$$

We prove the statement by using the continuous analog of Proposition 6.32, which will be established later, in Section 11.7.

It is sufficient to prove the statement for $p = \frac{d_A}{d_A+1}$. In this case, the depolarizing channel for $p > \frac{d_A}{d_A+1}$ can be represented as a mixture of this channel and a completely depolarizing channel $S \rightarrow \frac{I}{d_A} \text{Tr } S$, which are both entanglement-breaking. For $p = \frac{d_A}{d_A+1}$, we will prove

$$\Phi[S] \equiv \frac{1}{d_A+1} S + \frac{d_A}{d_A+1} \frac{I}{d_A} = d_A \int_{\Theta} |\theta\rangle\langle\theta| S |\theta\rangle\langle\theta| \nu(d\theta) \quad (6.56)$$

for all density operators S , which is a continuous analog of the Kraus representation (6.44), meaning that the channel is entanglement-breaking. We will use Lemma IV.4.1 from [107], according to which, for any continuous function F and a fixed $|\theta'\rangle$,

$$\int_{\Theta} F(|\langle\theta|\theta'\rangle|) \nu(d\theta) = - \int_0^1 F(r) d(1-r^2)^{d_A-1}. \quad (6.57)$$

It suffices to establish (6.56) for all $S = |\theta'\rangle\langle\theta'|$, $\theta' \in \Theta$. Consider the operator

$$\sigma = d_A \int_{\Theta} |\theta\rangle\langle\theta|\theta'\rangle\langle\theta'|\theta\rangle\langle\theta|\nu(d\theta). \quad (6.58)$$

It has unit trace, since

$$\text{Tr } \sigma = d_A \int_{\Theta} |\langle\theta|\theta'\rangle|^2 \nu(d\theta) = -d_A \int_0^1 r^2 d(1-r^2)^{d_A-1} = 1$$

by (6.57). From this, (6.55) follows by polarization.

Next, σ commutes with all unitaries leaving invariant $|\theta'\rangle$. Hence, it must have the form

$$\sigma = (1-p) |\theta'\rangle\langle\theta'| + p \frac{I}{d_A}.$$

To find p take $\langle\theta'|\sigma|\theta'\rangle$. We then obtain from (6.58)

$$(1-p) + \frac{p}{d_A} = d_A \int_{\Theta} |\langle\theta|\theta'\rangle|^4 \nu(d\theta) = -d_A \int_0^1 r^4 d(1-r^2)^{d_A-1}.$$

Computing the integral with the formula (6.57) we obtain the value $\frac{2}{d_A+1}$, whence $p = \frac{d_A}{d_A+1}$. \square

6.8 Qubit channels

Let us consider channels in the qubit space \mathcal{H}_2 which are affine maps of the Bloch ball in \mathbb{R}^3 , satisfying an additional restriction imposed by complete positivity.

Proposition 6.41. *An arbitrary linear, positive, trace-preserving map Φ of the space $\mathfrak{T}(\mathcal{H}_2)$ can be represented in the form*

$$\Phi[S] = U_2 \Lambda[U_1 S U_1^*] U_2^*, \quad (6.59)$$

where U_1, U_2 are unitary operators and Λ has the following canonical form, in the basis of Pauli matrices

$$\Lambda[I] = I + \sum_{\gamma=x,y,z} t_\gamma \sigma_\gamma, \quad \Lambda[\sigma_\gamma] = \lambda_\gamma \sigma_\gamma, \quad \gamma = x, y, z, \quad (6.60)$$

with λ_γ, t_γ real.

Proof. The restriction of Φ onto the state space of qubit is an affine map, hence it maps the state $S(\vec{a})$, given by (2.8) into the state $S(T\vec{a} + \vec{b})$, where T is a real 3×3 -matrix, \vec{b} is a vector in \mathbb{R}^3 . By using the polar decomposition for T and then the spectral decomposition for $|T|$, we have

$$T = O|T| = O_2 L O_1,$$

where O, O_1, O_2 are orthogonal matrices, $\det O_1 = \det O_2 = 1$ and $L = \text{diag}[\lambda_x, \lambda_y, \lambda_z]$ is diagonal, with the λ 's real but not necessarily nonnegative. Thus,

$$\Phi[S(\vec{a})] = S(O_2[L(O_1\vec{a}) + \vec{t}]), \quad (6.61)$$

where $\vec{t} = O_2^{-1}\vec{b}$.

The matrices O_1, O_2 describe rotations of the Bloch ball in \mathbb{R}^3 . Let us show that they correspond to reversible evolutions in \mathcal{H} , generated by unitary operators U in \mathcal{H} according to formula (6.2). It is sufficient to consider rotation O by an angle φ around the z -axis. In this case, a unit vector \vec{a} with the Euler angles θ, ϕ , is rotated to $O\vec{a}$, with the angles $\theta, \phi + \varphi$ so that the corresponding state vector (2.9) in \mathcal{H} is transformed into

$$|\psi(O\vec{a})\rangle = \begin{bmatrix} \cos \frac{\theta}{2} e^{-i(\phi+\varphi)/2} \\ \sin \frac{\theta}{2} e^{i(\phi+\varphi)/2} \end{bmatrix} = U |\psi(\vec{a})\rangle, \quad (6.62)$$

where $U = \text{diag}[e^{-i\varphi/2}, e^{i\varphi/2}]$ is a unitary matrix. Then

$$S(O\vec{a}) = US(\vec{a})U^* \quad (6.63)$$

for all \vec{a} .

By using (6.61) we obtain the required statement (6.59). \square

Exercise 6.42. Show that a rotation of the Bloch ball around the axis \vec{a} by the angle φ is implemented by the unitary operator

$$U = \exp\left[-\frac{i\varphi}{2}\sigma(\vec{a})\right] = \cos \frac{\varphi}{2} I - i \sin \frac{\varphi}{2} \sigma(\vec{a}).$$

Let us focus our attention on the maps of the form (6.60). Of course, complete positivity imposes nontrivial restrictions on the parameters λ_γ, t_γ [174]. Especially simple is the case where $t_\gamma \equiv 0$, when the map Φ , and hence Λ , is unital. In this case Λ is a contraction of the Bloch ball along the axes x, y, z , with coefficients $|\lambda_x|, |\lambda_y|, |\lambda_z|$, combined with reflections in case some of the numbers λ_γ are negative (for example, $\lambda_y < 0$ implies reflection with respect to the plane xz). Transposition of the density matrix $S(\vec{a})$, which is not completely positive, corresponds to Λ with the parameters $\lambda_x = 1, \lambda_y = -1, \lambda_z = 1$. It follows, in particular, that the relation of the type (6.63) also holds for orthogonal matrices O with $\det O = -1$, but with antiunitary operator U . Since any affine bijection of the Bloch ball is apparently implemented by an orthogonal matrix O , this implies Theorem 6.5 in the case of qubit.

Exercise 6.43. By using the multiplication rules (2.17), show that in the case $t_\gamma \equiv 0$

$$\Lambda[S] = \sum_{\gamma=0,x,y,z} \mu_\gamma \sigma_\gamma S \sigma_\gamma, \quad (6.64)$$

where

$$\mu_0 = \frac{1}{4} (1 + \lambda_x + \lambda_y + \lambda_z), \quad \mu_x = \frac{1}{4} (1 + \lambda_x - \lambda_y - \lambda_z),$$

$$\mu_y = \frac{1}{4} (1 - \lambda_x + \lambda_y - \lambda_z), \quad \mu_z = \frac{1}{4} (1 - \lambda_x - \lambda_y + \lambda_z).$$

Non-negativity of these numbers is the necessary and sufficient condition for the complete positivity of Λ and hence of Φ .

In the case $d = 2$, the “discrete Weyl operators” are nothing but the Pauli matrices, namely, with the convention $|e_0\rangle = |\uparrow\rangle, |e_1\rangle = |\downarrow\rangle$, we have

$$W_{01} = V = \sigma_z, \quad W_{10} = U = \sigma_x, \quad W_{11} = UV = -i\sigma_y, \quad (6.65)$$

and the “discrete CCR” are the multiplication rules (2.17).

Exercise 6.44. Show that qubit unital channels (6.64) are covariant with respect to the projective unitary representation of the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ defined by (6.65).

6.9 Notes and references

1. Proof of Theorem 6.5, going back to famous Wigner’s Theorem [220], follows the book of Davies (see Section 2.3 and Comments in [48]).

2. The notion of a completely positive map, introduced by Stinespring [202] in the more general context of C^* -algebras, was thoroughly studied in the paper of Arve-

son [9]. A useful survey of the properties of CP maps, including the proof of the inequality from Exercise 6.7, is in the paper of Störmer (pp. 85–106 in the volume [61]). In the finite-dimensional case, which has its own special features, CP maps were considered in detail in the work of Choi [41]. In these works one can find the conditions for extremality of CP maps.

This proof of Theorem 6.9 relies upon Gelfand–Naimark–Segal (GNS) construction, where from an object having a positivity property, one constructs a representation in a Hilbert space. The above proof, with a minor modification, works for CP maps of C^* -algebras in infinite dimensions. In the case of the algebra $\mathfrak{B}(\mathcal{H})$, Lemma 6.10 allows us to obtain the more special result (6.6).

3. The characteristic CP property of the dynamics of an open quantum system was observed in particular by Lindblad [148] and Holevo [92]. However, it was already implicit in the notion of “dynamical matrix” introduced by Sudarshan et al [206], as well as the representation (6.18). Equivalence of this representation with complete positivity was proved by Choi [41]. Later it was independently considered in great detail in the book of Kraus [139]. Some authors use the name *operator-sum representation* for (6.17), (6.18). The correspondence (6.26) was introduced by Jamiołkowski [123] for positive maps and by Choi [41] in the case of CP maps. Many interesting applications of this correspondence in the field of quantum information theory are considered by Verstraete and Verschelde [213].

Extensive literature is devoted to the Markovian dynamics of an open system described by a quantum dynamical semigroup, i.e. the semigroup of normalized positive or completely positive maps. A survey of quantum dynamical semigroups and closely related quantum stochastic processes can be found in the books [48], [99].

4. Q-c-q channels were introduced in the paper [98]. A detailed study of entanglement-breaking channels, including proof of Proposition 6.22, is given in the paper of M. Horodecki, Shor, and Ruskai [121].

Entanglement-binding channels were introduced in Horodeckis’ paper [118]), where examples are given of PPT channels that are not entanglement-breaking.

5. For detailed discussions of mathematical models for the quantum measurement process, see the books of Ludwig [152], Davies [48], Kraus [139], Holevo [99], the papers of Lindblad [146], and Ozawa [163].

6. The notion of complementary channel was introduced in the paper of Devetak and Shor [52]. The results in this section are due to Holevo [103], and to King, Matsumoto, Natanson, and Ruskai [132].

7. Discrete Weyl operators were used in the famous paper of Bennett, Brassard, Crépeau, Jozsa, Peres, and Wootters [19] on quantum teleportation. Concerning the covariance of the complementary channel (Exercise 6.38) see [103]. The Choi–Jamiołkowski state of the depolarizing channel is the *isotropic state* (a mixture of

chaotic and maximally entangled states) [170], and a usual way to prove Proposition 6.40, as well as its converse, is to study the separability of that state (see e.g. [128]).

8. Realistic examples of qubit channels are considered in detail in the book of Nielsen and Chuang [158]. The structure of general qubit channels is described in detail by Ruskai, Szarek, and E. Werner [174], see also Verstraete and Verschelde [213].

Chapter 7

Quantum entropy and information quantities

7.1 Quantum relative entropy

The classical relative entropy was considered in Section 4.2, where its monotonicity property and relation to the Shannon mutual information were established. The following definition introduces the noncommutative analog of the relative entropy, which plays an important role in quantum information theory. Many properties of the relative entropy are generalized to the quantum case, but some proofs become much more involved.

Definition 7.1. Let S, T be two density operators. If T is nondegenerate, the *quantum relative entropy* is defined by the formula

$$H(S; T) = -H(S) - \text{Tr } S \log T = \text{Tr } S(\log S - \log T). \quad (7.1)$$

The case where $\text{supp } S \subseteq \text{supp } T$ ($\text{supp } S$ denotes support of S , see Section 1.3) is reduced to the previous situation by considering the restrictions of the operators S, T to $\text{supp } T$, where T is nondegenerate. If however $\text{supp } S \not\subseteq \text{supp } T$, then, by definition, $H(S; T) = +\infty$.

Exercise 7.2. Let λ_j, μ_k be the eigenvalues and $|e_j\rangle, |h_k\rangle$ the respective eigenvectors of the density operators S, T . Prove that in this case,

$$H(S; T) = \sum_{j,k} |\langle e_j | h_k \rangle|^2 (\lambda_j \log \lambda_j - \lambda_j \log \mu_k). \quad (7.2)$$

Proposition 7.3.

$$H(S; T) \geq \frac{\log e}{2} \text{Tr } (S - T)^2 \geq 0, \quad (7.3)$$

with equalities if and only if $S = T$.

Proof. By using the formula

$$\eta(\lambda) - \eta(\mu) = (\lambda - \mu)\eta'(\mu) + \frac{1}{2}(\lambda - \mu)^2\eta''(\xi),$$

where ξ lies between λ and μ , we obtain for the function (4.2) the inequality

$$\lambda \log \lambda - \lambda \log \mu \geq \log e \left[(\lambda - \mu) + \frac{1}{2}(\lambda - \mu)^2 \right].$$

Substituting this into (7.2), we get the result. \square

Corollary 7.4 (Subadditivity of Quantum Entropy). *Let S_{12} be a density operator in $\mathcal{H}_1 \otimes \mathcal{H}_2$, with partial states $S_1 = \text{Tr}_{\mathcal{H}_2} S_{12}$ and $S_2 = \text{Tr}_{\mathcal{H}_1} S_{12}$. Then*

$$H(S_{12}) \leq H(S_1) + H(S_2), \quad (7.4)$$

with equality if and only if $S_{12} = S_1 \otimes S_2$.

Proof. This follows from the equality

$$H(S_1) + H(S_2) - H(S_{12}) = H(S_{12}; S_1 \otimes S_2)$$

and the inequality (7.3). \square

Similar to the ordinary entropy (see Exercise 5.3), the relative entropy has the following elementary properties, to be proved in the following exercise:

Exercise 7.5.

- i. *Isometry invariance:* $H(VSV^*; VS'V^*) = H(S; S')$ for an arbitrary operator V isometric on the supports of S, S' ;
- ii. *Additivity:* $H(S_1 \otimes S_2; S'_1 \otimes S'_2) = H(S_1; S'_1) + H(S_2; S'_2)$.

7.2 Monotonicity of the relative entropy

The following key property is the noncommutative analog of the inequality (4.18).

Theorem 7.6 (Lindblad [148]). *For arbitrary density operators S, T in \mathcal{H}_1 , and arbitrary channel $\Phi : \mathfrak{T}(\mathcal{H}_1) \rightarrow \mathfrak{T}(\mathcal{H}_2)$*

$$H(\Phi[S]; \Phi[T]) \leq H(S; T). \quad (7.5)$$

This important result follows from Theorem 7.7, to be proved below, which establishes a similar monotonicity property for a whole class of functions of a pair of quantum states.

Consider the Hilbert space $L^2(\mathcal{H})$ of linear operators in \mathcal{H} , with the inner product

$$(X, Y) = \text{Tr } X^* Y.$$

Let S, T be nondegenerate operators. In $L^2(\mathcal{H})$, we introduce the operator L_T of left multiplication by T and the operator R_S of right multiplication by S , namely, $L_T X = TX$, $R_S X = XS$. They are positive commuting Hermitian operators in the Hilbert space $L^2(\mathcal{H})$.

The *relative g -entropy* of operators S, T is defined by the relation

$$H_g(S; T) = (S^{1/2}, g(L_T R_S^{-1}) S^{1/2}), \quad (7.6)$$

for any function $g \in \mathcal{G}$, where \mathcal{G} is the class of functions of the form

$$g(w) = a(w-1) + b(w-1)^2 + \int_0^\infty \frac{(w-1)^2}{w+s} d\nu(s), \quad (7.7)$$

with a real, $b \geq 0$, and ν a positive measure on $[0, \infty)$ such that $\int_0^\infty \frac{1}{s+1} d\nu(s) < \infty$ (concerning the inner characterization of \mathcal{G} as the class of operator-convex functions such that $g(1) = 0$ see Section 7.8). For the proof of the following theorem it is only important that the function $g(w)$ admits such a representation.

Theorem 7.7 (Petz [167]). *For arbitrary $g \in \mathcal{G}$, the relative g -entropy $H_g(S; T)$ has the monotonicity property*

$$H_g(\Phi[S]; \Phi[T]) \leq H_g(S; T), \quad (7.8)$$

where Φ is an arbitrary channel.

Theorem 7.6 then follows from the two observations,

- i. The function $g(w) = -\log w$ belongs to the class \mathcal{G} . Indeed, by performing integration by parts, one can check that

$$-\ln w = \int_0^\infty \left[\frac{1}{w+s} - \frac{1}{1+s} \right] ds = -(w-1) + \int_0^\infty \frac{(w-1)^2}{(w+s)(s+1)^2} ds.$$

- ii.

$$H_{-\log}(S; T) = H(S; T).$$

We have

$$-\log(L_T R_S^{-1})X = (\log R_S)X - (\log L_T)X = X(\log S) - (\log T)X$$

because of the commutativity of the operators L_T, R_S , whence we obtain $-\log(L_T R_S^{-1})S^{1/2} = (\log S - \log T)S^{1/2}$, and the result follows from (7.6).

Proof of Theorem 7.7.

Lemma 7.8. *For an arbitrary function $g \in \mathcal{G}$,*

$$H_g(S; T) = b \operatorname{Tr} (T - S) S^{-1} (T - S) + \int_0^\infty \operatorname{Tr} (T - S) (L_T + s R_S)^{-1} (T - S) dv(s) \quad (7.9)$$

Proof. By introducing the notation $\Delta_{T,S} = L_T R_S^{-1}$, we see that

$$(\Delta_{T,S} - I) S^{1/2} = (T - S) S^{-1/2} = R_{S^{-1/2}} (T - S), \quad (7.10)$$

hence,

$$H_{w-1}(S; T) = \operatorname{Tr} S^{1/2} (T - S) S^{-1/2} = 0, \quad (7.11)$$

so that the linear term in (7.7) does not contribute to (7.9). For $g(w) = \frac{(w-1)^2}{(w+s)}$, $s > 0$, we find, by using (7.10),

$$\begin{aligned} H_g(S; T) &= ((\Delta_{T,S} - I) S^{1/2}, (\Delta_{T,S} + sI)^{-1} (\Delta_{T,S} - I) S^{1/2}) \\ &= \operatorname{Tr} [(T - S) (\Delta_{T,S} + sI)^{-1} R_{S^{-1}} (T - S)] \\ &= \operatorname{Tr} (T - S) (L_T + s R_S)^{-1} (T - S). \end{aligned} \quad (7.12)$$

For $s = 0$ we get

$$H_{(w-1)^2/w}(S; T) = \operatorname{Tr} (T - S) T^{-1} (T - S) = H_{(w-1)^2}(T; S). \quad (7.13)$$

By substituting this into (7.7), we get (7.9). \square

If we neglect the first term in (7.7) (which does not contribute to $H_g(S; T)$), the functions (7.7) are “continual convex combinations” of the functions $(w-1)^2$ and $\frac{(w-1)^2}{w+s}$, $s \geq 0$. Thus, it is sufficient to prove monotonicity of the relative g -entropy for such functions g , i.e. for (7.12).

The proof of Theorem 7.7 now follows from the integral representation (7.9) and the following

Lemma 7.9. *For arbitrary channel Φ , $s \geq 0$, and an operator A*

$$\operatorname{Tr} A^* (L_T + s R_S)^{-1} A \geq \operatorname{Tr} \Phi[A^*] (L_{\Phi[T]} + s R_{\Phi[S]})^{-1} \Phi[A]. \quad (7.14)$$

Proof. Note that positive maps respect Hermitian conjugation, so that $\Phi[X^*] = \Phi[X]^*$, and similarly for Φ^* .

Since L_T, R_S are positive operators in $L^2(\mathcal{H})$, the operator $L_T + s R_S$ is also positive. Set $X = (L_T + s R_S)^{-1/2} A - (L_T + s R_S)^{1/2} \Phi^*[B]$, where $B = (L_{\Phi[T]} + s R_{\Phi[S]})^{-1} \Phi[A]$. In this case, $\operatorname{Tr} X^* X \geq 0$, so that

$$\begin{aligned} &\operatorname{Tr} A^* (L_T + s R_S)^{-1} A - \operatorname{Tr} A^* \Phi^*[B] \\ &\quad - \operatorname{Tr} \Phi^*[B^*] A + \operatorname{Tr} \Phi^*[B^*] (L_T + s R_S) \Phi^*[B] \geq 0. \end{aligned} \quad (7.15)$$

We have

$$\mathrm{Tr} A^* \Phi^*[B] + \mathrm{Tr} \Phi^*[B^*]A = 2\mathrm{Tr} \Phi[A^*](L_{\Phi[T]} + sR_{\Phi[S]})^{-1} \Phi[A],$$

therefore, in order to prove the lemma, it is sufficient to show that the last term in (7.15) is less than or equal to the right-hand side of (7.14). Thus,

$$\begin{aligned} \mathrm{Tr} \Phi^*[B^*](L_T + sR_S)\Phi^*[B] &= \mathrm{Tr} [s\Phi^*[B^*]\Phi^*[B]S + \Phi^*[B^*]T\Phi^*[B]] \\ &= \mathrm{Tr} [s\Phi^*[B^*]\Phi^*[B]S + \Phi^*[B]\Phi^*[B^*]T] \\ &\leq \mathrm{Tr} [s\Phi^*[B^*B]S + \Phi^*[BB^*]T], \end{aligned}$$

where the inequality follows from the positivity of S, T and the Kadison inequality (6.5), which in the case of a channel Φ takes the form

$$\Phi^*[B^*]\Phi^*[B] \leq \Phi^*[B^*B].$$

Then by using the relation $\mathrm{Tr} \Phi^*[B^*B]S = \mathrm{Tr} B^*B\Phi[S]$, we find

$$\begin{aligned} \mathrm{Tr} \Phi^*[B^*](L_T + sR_S)\Phi^*[B] &\leq \mathrm{Tr} (sB^*B\Phi[S] + BB^*\Phi[T]) \\ &= \mathrm{Tr} B^*(sB\Phi[S] + \Phi[T]B) \\ &= \mathrm{Tr} B^*(L_{\Phi[T]} + sR_{\Phi[S]})(B) = \mathrm{Tr} B^*\Phi[A] \\ &= \mathrm{Tr} \Phi[A^*](L_{\Phi[T]} + sR_{\Phi[S]})^{-1} \Phi[A]. \quad \square \end{aligned}$$

This proves Theorem 7.7 in the case of nondegenerate S, T . In the general case, either $H(S; T) = +\infty$ and the inequality (7.5) is trivial, or $\mathrm{supp} S \subseteq \mathrm{supp} T$, in which case we can assume that $\mathrm{supp} T = \mathcal{H}$, i.e. the operator T is nondegenerate and for $H(S; T)$ the formula (7.1) holds. Approximating an arbitrary operator S by nondegenerate operators and using the continuity of the entropy (see Section 7.4), proves the statement of the theorem in the general case. \square

We now consider a number of useful corollaries of the Monotonicity Theorem 7.6.

Corollary 7.10 (Generalized H-theorem). *Let Φ be a bistochastic (unital) channel in $\mathcal{H} = \mathcal{H}_1 = \mathcal{H}_2$. Then, for any state S ,*

$$H(\Phi[S]) \geq H(S). \quad (7.16)$$

Proof. From (7.1) it follows that

$$H(S) = \log d - H(S; \bar{S}), \quad (7.17)$$

where $\bar{S} = \frac{1}{d}I$ is the chaotic state, hence

$$H(\Phi[S]) = \log d - H(\Phi[S]; \Phi[\bar{S}])$$

because $\Phi[\bar{S}] = \bar{S}$. The corollary now follows by virtue of the Monotonicity Theorem 7.6. \square

Corollary 7.11. *The quantity (5.8) is monotone under the action of the arbitrary channel Φ :*

$$\chi(\{\pi_x\}; \Phi[\{S_x\}]) \leq \chi(\{\pi_x\}; \{S_x\}). \quad (7.18)$$

Proof. By introducing the average state $\bar{S}_\pi = \sum \pi_x S_x$, we obtain the important identity generalizing (4.21)

$$\chi(\{\pi_x\}; \{S_x\}) \equiv H(\bar{S}_\pi) - \sum \pi_x H(S_x) = \sum \pi_x H(S_x; \bar{S}_\pi). \quad (7.19)$$

Therefore (7.18) follows directly from (7.5). \square

In particular, Theorem 7.6 implies the bound (5.16) on the Shannon information. Take an arbitrary basis $\{e_y\}$ in \mathcal{H} and consider the q-c channel

$$\Psi[S] = \sum_y \text{Tr } S M_y |e_y\rangle\langle e_y|,$$

corresponding to the measurement of the observable $M = \{M_y\}$, then

$$\begin{aligned} \Psi[S_x] &= \sum_y p_M(y|x) |e_y\rangle\langle e_y|; \\ \Psi[\bar{S}_\pi] &= \sum_y \left(\sum_x \pi_x p_M(y|x) \right) |e_y\rangle\langle e_y|, \end{aligned}$$

where $p_M(y|x) = \text{Tr } S_x M_y$, are the density operators that are diagonal in the basis $\{e_y\}$. Therefore, the quantum relative entropy at the output of the channel Ψ becomes the classical relative entropy, and relation (4.21) implies

$$\sum_x \pi_x H(\Psi[S_x]; \Psi[\bar{S}_\pi]) = \mathcal{I}_1(\pi, M).$$

By using the relation (7.18) for the channel Ψ we obtain the required inequality.

A useful quantity, motivated by Corollary 7.10, is the *entropy gain*

$$G_\Phi[S] = H(\Phi[S]) - H(S). \quad (7.20)$$

Corollary 7.12. *The entropy gain is convex as a function of S and is superadditive:*

$$G_{\Phi_1 \otimes \Phi_2}[S_{12}] \geq G_{\Phi_1}[S_1] + G_{\Phi_2}[S_2]. \quad (7.21)$$

Proof. Taking into account the first identity in (7.19), the convexity of $G_\Phi[S]$ is just a reformulation of relation (7.18). The superadditivity (7.21) is equivalent to the relation

$$H((\Phi_1 \otimes \Phi_2)[S_{12}]; (\Phi_1 \otimes \Phi_2)[S_1 \otimes S_2]) \leq H(S_{12}; S_1 \otimes S_2)$$

which follows from the Monotonicity Theorem 7.6. \square

7.3 Strong subadditivity of the quantum entropy

Theorem 7.13. *The following properties are equivalent:*

- i. *monotonicity of the quantum relative entropy*
- ii. *strong subadditivity of the quantum entropy: Let S_{123} be a density operator in $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. In this case, with the obvious notations for the partial states,*

$$H(S_{123}) + H(S_2) \leq H(S_{12}) + H(S_{23}). \quad (7.22)$$

- iii. *joint convexity of the quantum relative entropy $H(S; T)$ with respect to the arguments S, T*

Proof.

- i. \Rightarrow ii. Denoting by $\bar{S}_\alpha = (\dim \mathcal{H}_\alpha)^{-1} I_\alpha$ the chaotic state in \mathcal{H}_α , we have

$$\begin{aligned} H(S_{123}; \bar{S}_{123}) &= H(S_{123}; S_{12} \otimes \bar{S}_3) + H(S_{12}; \bar{S}_{12}), \\ H(S_{23}; \bar{S}_{23}) &= H(S_{23}; S_2 \otimes \bar{S}_3) + H(S_2; \bar{S}_2). \end{aligned}$$

Performing subtraction and using (7.17), we get

$$H(S_{12}) + H(S_{23}) - H(S_{123}) - H(S_2) = H(S_{123}; S_{12} \otimes \bar{S}_3) - H(S_{23}; S_2 \otimes \bar{S}_3).$$

Consider the partial trace channel $\Phi[S_{123}] = S_{23}$. Applying Theorem 7.6 to it, we deduce the nonnegativity of the right hand side, i.e. the strong subadditivity.

- ii. \Rightarrow iii. Consider the density operator in $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ of the form

$$S_{123} = \sum_{kl} |e_k\rangle\langle e_k| \otimes A_{kl} \otimes |h_l\rangle\langle h_l|,$$

where $\{e_k\}$ is an orthonormal basis in \mathcal{H}_1 , $\{h_l\}$ is an orthonormal basis in \mathcal{H}_3 , and A_{kl} are positive operators in \mathcal{H}_2 , which become density operators with an appropriate normalization. In this case,

$$S_{12} = \sum_k |e_k\rangle\langle e_k| \otimes \sum_l A_{kl}, \quad S_{23} = \sum_{kl} A_{kl} \otimes |h_l\rangle\langle h_l|, \quad S_2 = \sum_{kl} A_{kl},$$

and the strong subadditivity implies

$$-\sum_{kl} H(A_{kl}) + \sum_k H\left(\sum_l A_{kl}\right) + \sum_l H\left(\sum_k A_{kl}\right) - H\left(\sum_{kl} A_{kl}\right) \geq 0,$$

where we denoted $H(A) = -\text{Tr} A \log A$ for a positive operator A . In particular, for arbitrary positive operators A_1, A_2, B_1, B_2 ,

$$\begin{aligned} H(A_1 + A_2 + B_1 + B_2) \\ - H(A_1 + A_2) - H(B_1 + B_2) \leq H(A_1 + B_1) - H(A_1) - H(B_1) \\ + H(A_2 + B_2) - H(A_2) - H(B_2), \end{aligned}$$

which is equivalent to the joint convexity of the function

$$\Delta(A, B) = H(A + B) - H(A) - H(B)$$

with respect to A, B . On the other hand, the following lemma implies that $H(A; B)$ is convex as a supremum of a family of convex functions.

Lemma 7.14. *For any two density operators A, B*

$$H(A; B) = \sup_{\lambda > 0} \lambda^{-1} \chi_\lambda(A; B), \quad (7.23)$$

where

$$\begin{aligned} \chi_\lambda(A; B) &= H(\lambda A + (1 - \lambda)B) - \lambda H(A) - (1 - \lambda)H(B) \\ &= \Delta(\lambda A, (1 - \lambda)B) + h(\lambda). \end{aligned}$$

Proof. From the proof above, it follows that the function $\chi_\lambda(A; B)$ is convex with respect to A, B . As a function of $\lambda \in [0, 1]$ it is concave and equal to zero for $\lambda = 0, 1$ (this was proved in Section 5.3). Hence, for fixed A, B

$$\sup_{\lambda > 0} \lambda^{-1} \chi_\lambda(A; B) = \lim_{\lambda \rightarrow +0} \lambda^{-1} \chi_\lambda(A; B) = \left. \frac{d}{d\lambda} \chi_\lambda(A; B) \right|_{\lambda=0}.$$

Exercise 7.15. Let F be a continuously differentiable function on $(0, +\infty)$. Show that for any two positive operators A, B the following holds:

$$\frac{d}{d\lambda} \text{Tr} F(\lambda A + (1 - \lambda)B) = \text{Tr} F'(\lambda A + (1 - \lambda)B)(A - B). \quad (7.24)$$

Hint: prove (7.24) for the case where $F(x) = x^k$, implying the result for polynomials, and then use the Weierstrass uniform approximation theorem.

Now let $\text{supp } A \subseteq \text{supp } B$, so that $H(A; B) < +\infty$. Without loss of generality we assume that the operator B is nondegenerate. Computing the derivative of

$$H(\lambda A + (1 - \lambda)B) = \text{Tr} \eta(\lambda A + (1 - \lambda)B)$$

by formula (7.24) we obtain

$$\left. \frac{d}{d\lambda} \chi_\lambda(A; B) \right|_{\lambda=0} = \text{Tr } A(\log A - \log B) + (\log e) \text{Tr } (B - A) = H(A; B),$$

from which (7.23) follows. \square

■ **Exercise 7.16.** Prove the statement (7.23) for the case $H(A; B) = +\infty$.

iii. \Rightarrow i. Consider the density operator S_{12} in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and let $S_1 = \text{Tr}_2 S_{12}$. From the property (6.53) of the discrete Weyl operators, we obtain

$$S_1 \otimes \bar{S}_2 = \frac{1}{d^2} \sum_{\alpha, \beta=0}^{d-1} (I_1 \otimes W_{\alpha\beta}) S_{12} (I_1 \otimes W_{\alpha\beta})^*.$$

The joint convexity of the relative entropy and the properties from Exercise 7.5 then imply

$$H(S_1; S_1') \leq H(S_{12}; S_{12}'),$$

which is monotonicity with respect to the partial trace channel. But, according to Theorem 6.9, every channel can be represented as a concatenation of an isometric embedding (preserving the relative entropy) and a partial trace. This proves i. \square

By using (7.17) and the convexity of the relative entropy we also obtain

Corollary 7.17. *The quantum entropy is concave: for arbitrary states S_j and probability distribution $\{p_j\}$,*

$$H\left(\sum_j p_j S_j\right) \geq \sum_j p_j H(S_j).$$

■ **Exercise 7.18.** Find a direct proof of the concavity, using the spectral decomposition of the operator $\sum_j p_j S_j$ and the concavity of the function $\eta(t)$.

7.4 Continuity properties

In the finite-dimensional case that we consider here, the quantum entropy is a continuous function of a state. Indeed, the function $\eta(x) = -x \log x$ is (uniformly) continuous on the segment $[0, 1]$. Hence, $S_n \rightarrow S$ implies $\|\eta(S_n) - \eta(S)\| \rightarrow 0$, where $\|\cdot\|$ is the operator norm. Therefore,

$$|H(S_n) - H(S)| = |\text{Tr } (\eta(S_n) - \eta(S))| \leq d \|\eta(S_n) - \eta(S)\| \rightarrow 0.$$

A more precise estimate, which we will need later, is given by the following

Lemma 7.19 (Fannes [58]). *Let S_1, S_2 be two density operators in d -dimensional Hilbert space, such that $\|S_1 - S_2\|_1 \leq \frac{1}{e}$. In this case,*

$$|H(S_1) - H(S_2)| \leq \log d \cdot \|S_1 - S_2\|_1 + \eta(\|S_1 - S_2\|_1). \quad (7.25)$$

Since the function $\eta(x)$ is monotonously increasing for $x \in [0, \frac{1}{e}]$, the lemma implies the estimate

$$|H(S_1) - H(S_2)| \leq \log d \cdot \|S_1 - S_2\|_1 + \frac{\log e}{e}. \quad (7.26)$$

Proof. Let $\lambda_1 \geq \lambda_2 \geq \dots$ ($\mu_1 \geq \mu_2 \geq \dots$) be the eigenvalues of S_1 (correspondingly, S_2).

Lemma 7.20. *Denoting $\Delta_i = |\lambda_i - \mu_i|$, one has*

$$\Delta = \sum_{i=1}^d \Delta_i \leq \|S_1 - S_2\|_1. \quad (7.27)$$

Proof. Let $v_1 \geq v_2 \geq \dots$ be the eigenvalues of the operator

$$A = S_1 + [S_1 - S_2]_- = S_2 + [S_1 - S_2]_+.$$

Since $A \geq S_1, S_2$, we have $v_i \geq \lambda_i, \mu_i$, so that

$$|\lambda_i - \mu_i| \leq 2v_i - \lambda_i - \mu_i.$$

Therefore,

$$\begin{aligned} \Delta &\leq \sum_{i=1}^d (2v_i - \lambda_i - \mu_i) = \text{Tr}(2A - S_1 - S_2) = \text{Tr}[S_1 - S_2]_+ + \text{Tr}[S_1 - S_2]_- \\ &= \|S_1 - S_2\|_1. \end{aligned} \quad \square$$

Exercise 7.21. Prove the inequality

$$|\eta(y) - \eta(x)| \leq \eta(y - x), \quad 0 \leq x \leq y \leq 1.$$

Combining these statements, we obtain

$$\begin{aligned} |H(S_1) - H(S_2)| &\leq \sum_{i=1}^d |\eta(\lambda_i) - \eta(\mu_i)| \leq \sum_{i=1}^d \eta(\Delta_i) \\ &= \Delta \sum_{i=1}^d \eta(\Delta_i / \Delta) + \eta(\Delta) \leq \Delta \log d + \eta(\Delta). \end{aligned}$$

Taking into account the monotonicity of $\eta(x)$, and combining this with (7.27) produces (7.25). \square

Exercise 7.22. Show that the relative entropy is lower semicontinuous in the following sense: if $S_n \rightarrow S$ and $S'_n \rightarrow S'$, then

$$\liminf_{n \rightarrow \infty} H(S_n; S'_n) \geq H(S; S').$$

Hint: use the continuity of the entropy, representation (7.23), and the fact that the supremum of a family of continuous functions is lower semicontinuous.

As we shall see later, in Chapter 11, this property survives for both the entropy and for the relative entropy in the infinite-dimensional Hilbert space, while the global continuity of the entropy no longer holds.

7.5 Information correlation, entanglement of formation and conditional entropy

In the classical case, the amount of information transmitted by the channel $X \rightarrow Y$ is measured by the quantity $I(X; Y) = H(X) + H(Y) - H(XY)$. In quantum statistics there is no general analog of the quantity $H(XY)$, as the joint distribution of observables exists only in special cases. One such case is that of the composite system.

Let S_{12} be a state of a bipartite quantum system in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$. In this situation, an obvious quantum analogue of the Shannon mutual information is

$$I(1; 2) = H(S_1) + H(S_2) - H(S_{12}) = H(S_{12}; S_1 \otimes S_2). \quad (7.28)$$

We call this *information correlation*. From Corollary 7.4, it follows that $I(1; 2) \geq 0$ and $I(1; 2) = 0$ if and only if $S_{12} = S_1 \otimes S_2$.

Take a bipartite system and perform the measurements of observables M_1, M_2 in the corresponding subsystems. Their joint distribution exists and is given by the formula $p(x, y) = \text{Tr } S_{12}(M_{1x} \otimes M_{2y})$. Let $I(M_1; M_2)$ denote the Shannon mutual information between the outcomes of these measurements.

Exercise 7.23. Show that, for a pure state, $S_{12} = |\psi\rangle\langle\psi|$

$$I(1; 2) = 2 \max_{M_1, M_2} I(M_1, M_2). \quad (7.29)$$

In the case of the pure state S_{12} we have $H(S_{12}) = 0$, and Theorem 3.10 implies that

$$H(S_1) = H(S_2) \quad (7.30)$$

and

$$I(1; 2) = H(S_1) + H(S_2) = 2H(S_1). \quad (7.31)$$

The quantity $H(S_1) = H(S_2)$ then is a natural *measure of entanglement* of a pure bipartite state S_{12} . It is equal to zero for unentangled states and takes the maximal value $\log d$ for the state (3.7), which explains the name “maximally entangled state”.

The question of measures of entanglement for mixed states is substantially more complicated than that for pure states. Unlike the case of pure states there are many different measures of entanglement for mixed states. Of main interest for us will be *entanglement of formation*, which is defined for an arbitrary state S_{12} in $\mathcal{H}_1 \otimes \mathcal{H}_2$ as

$$E_F(S_{12}) = \inf \sum_j \pi_j H(\text{Tr}_2 S_{12}^j), \quad (7.32)$$

where the infimum is taken over all possible decompositions of the density operator

$$S_{12} = \sum_j \pi_j S_{12}^j$$

as convex combinations of some density operators S_{12}^j (in terms of convex analysis, $E_F(S_{12})$ is the *convex hull* of the continuous function $S_{12} \rightarrow H(S_{12})$).

Proposition 7.24. *The function $S_{12} \rightarrow E_F(S_{12})$ is continuous, convex, and the infimum in (7.32) is attained on a convex combination of no more than $(d_1 d_2)^2$ pure states S_{12}^j . Moreover, the duality relation holds:*

$$E_F(S_{12}) = \max_{A_{12}} \text{Tr } S_{12} A_{12}, \quad (7.33)$$

where the maximization is over the Hermitian operators A_{12} in $\mathcal{H}_1 \otimes \mathcal{H}_2$, satisfying

$$\text{Tr } T_{12} A_{12} \leq H(\text{Tr}_2 T_{12}), \quad (7.34)$$

for all (pure) states T_{12} in $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Proof. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$, $S = S_{12}$ and $f(S) = H(\text{Tr}_2 S_{12})$. In this case, f is a continuous concave function on the compact convex set of quantum states $\mathfrak{S} = \mathfrak{S}(\mathcal{H})$. The proof below uses very few of the other special properties of f and \mathfrak{S} , and could be entirely based on references to the convex duality theory. However, we here provide a self-contained proof.

A finite collection of states with the corresponding probabilities is usually called an *ensemble*¹. For our proof it is useful to consider “continual” generalization of ensembles, described by the probability measures $\pi(dS)$ on the set $\mathfrak{S}(\mathcal{H})$ (the concept of a generalized ensemble will be systematically explored in Ch. 11). Now, the usual ensembles correspond to finitely supported measures. By using general facts from measure theory (see e.g. [164]) one can show that the set $\mathcal{P}(\mathfrak{S}(\mathcal{H}))$ of all probability

¹ This notion is different from the “statistical ensemble” in Section 2.1.2, which means a statistical sample of identically prepared systems.

measures on $\mathfrak{S}(\mathcal{H})$ is itself compact with respect to the topology of weak convergence, which is defined as the convergence of integrals of all bounded continuous functions on $\mathfrak{S}(\mathcal{H})$.

Consider the functional

$$F(\pi) = \int_{\mathfrak{S}} f(T) \pi(dT), \quad (7.35)$$

on $\mathcal{P}(\mathfrak{S}(\mathcal{H}))$, which for finitely supported π coincides with the minimized expression in (7.32). From the definition, it is a continuous, affine functional. We are interested in the minimization of this functional, on the closed convex subset \mathcal{P}_S of probability measures π with the given barycenter

$$S = \int_{\mathfrak{S}(\mathcal{H})} T \pi(dT). \quad (7.36)$$

The continuous functional $F(\pi)$ attains its minimum on the compact set \mathcal{P}_S . The concavity of f then implies that we can choose the minimizing measure π to be supported by the pure states, since we can always perform decompositions of all density operators $T = T_{12}$ into pure states, without changing the barycenter and without increasing the value $F(\pi)$.

Let us now show that such an extreme π is supported by no more than $(d_1 d_2)^2$ states. Note that the states supporting π are linearly independent in the sense that if

$$\int_{\mathfrak{S}(\mathcal{H})} c(T) T \pi(dT) = 0 \quad (7.37)$$

for some real bounded measurable function $c(T)$, then $c(T) = 0$ almost everywhere, with respect to π . Indeed, it follows from (7.37) that $\pi = \frac{1}{2}\pi_+ + \frac{1}{2}\pi_-$, where

$$\pi_{\pm}(dT) = (1 \pm \varepsilon c(T)) \pi(dT)$$

for sufficiently small ε , so that extremality implies $\pi_{\pm}(dT) = \pi(dT)$, and hence $c(T) = 0 \pmod{\pi}$, i.e. almost everywhere with respect to the measure π . By taking the matrix elements of (7.37), we see that the real Hilbert space $L_{\mathbb{R}}^2(\pi)$ is spanned by $(d_1 d_2)^2$ functions

$$T \rightarrow \Re \langle e_j | T | e_k \rangle; \quad 1 \leq j \leq k \leq d_1 d_2,$$

$$T \rightarrow \Im \langle e_j | T | e_k \rangle; \quad 1 \leq j < k \leq d_1 d_2.$$

Indeed, any function $c(T)$ that is orthogonal to all these functions in $L_{\mathbb{R}}^2(\pi)$, must be zero $\pmod{\pi}$. The rest follows from the following exercise.

Exercise 7.25. Show that if $\dim L_{\mathbb{R}}^2(\pi) = n$, then π is finitely supported by n points.

The inequality \geq in the duality relation follows from integrating the inequality $f(T) \geq \text{Tr } T\Lambda$ with respect to an arbitrary probability measure π that satisfies the constraint (7.36). In the proof of the reverse inequality we can consider only those measures which have a fixed (but arbitrary) finite support, containing the support of the optimal (minimizing) measure π^0 . In this case, the minimization in (7.32) becomes a finite dimensional problem with a finite number of linear constraints, defined by (7.36). Note that these constraints also imply the constraint expressed by the normalization of a probability measure. Applying the elementary Lagrange method, we then obtain that there exists a Hermitian operator Λ such that π^0 minimizes the functional

$$F(\pi) - \text{Tr} \left(\int_{\mathfrak{E}} T\pi(dT) \right) \Lambda = \int_{\mathfrak{E}} [f(T) - \text{Tr } T\Lambda] \pi(dT) \quad (7.38)$$

over all (non-normalized) positive measures π with the given finite support. Here, the Hermitian operator Λ simply describes the collection of real Lagrange multipliers for the constraints (7.36). It follows that

$$f(T) - \text{Tr } T\Lambda \geq 0 \quad (7.39)$$

$$f(T) - \text{Tr } T\Lambda = 0 \pmod{\pi^0}. \quad (7.40)$$

The inequality (7.39) follows from the fact that otherwise the infimum of the functional (7.38) would be $-\infty$, while the equality (7.40) guarantees that the infimum, equal to 0, is attained. By integrating the second equality with respect to π^0 , and taking into account (7.36), we obtain $\text{Tr } S\Lambda = \int_{\mathfrak{E}} f(T)\pi^0(dS)$ which, together with inequality (7.39), implies that Λ is the solution of the dual problem, and (7.33) holds with the common value $\text{Tr } S\Lambda$.

The duality relation implies that the function $E_F(S)$ is lower semicontinuous, as the maximum of the family of continuous (affine) functions $S \rightarrow \text{Tr } S\Lambda$. It now suffices to show that it is upper semicontinuous, i.e.

$$\limsup_{n \rightarrow \infty} E_F(S^{(n)}) \leq E_F(S),$$

for any sequence of states $S^{(n)} \rightarrow S$. In this proof we follow Shirokov [185]. Let

$$E_F(S) = \sum_{j=1}^N \pi_j^0 H(\text{Tr } \mathcal{H}_2 S_j), \quad (7.41)$$

where $S = \sum_{j=1}^N \pi_j^0 S_j$. Now, if S is nondegenerate, put

$$S^{(n)} = \sum_{j=1}^N \pi_j^{(n)} S_j^{(n)},$$

where

$$\begin{aligned}\pi_j^{(n)} &= t_j^{(n)} \pi_j^0, \\ S_j^{(n)} &= (t_j^{(n)})^{-1} (S^{(n)})^{1/2} S^{-1/2} S_j S^{-1/2} (S^{(n)})^{1/2},\end{aligned}$$

and

$$t_j^{(n)} = \text{Tr} (S^{(n)})^{1/2} S^{-1/2} S_j S^{-1/2} (S^{(n)})^{1/2}.$$

These expressions also make sense for degenerate S , if by $S^{-1/2}$ we denote the generalized inverse of $S^{1/2}$, i.e. the operator that is equal to the inverse on the support of S and zero on its orthogonal complement (see [185] for detail). Now,

$$E_F(S^{(n)}) \leq \sum_{j=1}^N \pi_j^{(n)} H\left(\text{Tr} \mathcal{H}_2 S_j^{(n)}\right),$$

and the right-hand side tends to (7.41). Therefore

$$\limsup_{n \rightarrow \infty} E_F(S^{(n)}) \leq E_F(S),$$

that is E_F is upper semicontinuous, and hence continuous. \square

Let us come back to the case of the pure state S_{12} . Note that in classical statistics, partial states of a pure state (marginal distributions of a degenerate distribution) are again pure, and purification has no classical counterpart. Related to this fact is the following unusual property of the quantum analog of conditional entropy. In the classical case, the conditional entropy is always nonnegative

$$H(X|Y) = H(XY) - H(Y) = \sum_y p_y H(X|Y = y) \geq 0. \quad (7.42)$$

Defining the *quantum conditional entropy* as

$$H(1|2) = H(S_{12}) - H(S_2),$$

we see that it is negative if S_{12} is a pure entangled state. In other words, unlike the classical entropy, the quantum entropy is not monotone with respect to enlargement of the system: $H(S_1) \not\leq H(S_{12})$. Nevertheless, similar to the classical case, the following properties hold

- i. Monotonicity of the conditional entropy: $H(1|23) \leq H(1|2)$;
- ii. $H(1|2) + H(1|3) \geq 0$.

Exercise 7.26. Show that these properties are reformulations of the strong subadditivity of the quantum entropy (7.22). Hint: use purification of the total state.

Another useful property of the quantum conditional entropy is:

Corollary 7.27. *The conditional entropy $H(1|2)$ is a concave function of the state S_{12} . It is subadditive in the following sense,*

$$H(13|24) \leq H(1|2) + H(3|4). \quad (7.43)$$

The statements follow from Corollary 7.12 by observing that the conditional entropy is equal to minus the entropy gain that corresponds to the partial trace channel: $H(1|2) = -G_{\text{Tr}_1}(S_{12})$.

These properties make the conditional entropy a useful tool in quantum information theory.

7.6 Entropy exchange

Let Φ be a channel from the input space \mathcal{H}_A to the output space \mathcal{H}_B and let $S = S_A$ be a density operator in \mathcal{H}_A denoting the input state. By Theorem 6.9, the channel admits the Stinespring representation

$$\Phi[S_A] = \text{Tr}_E V S_A V^*, \quad (7.44)$$

where V is an isometric operator from \mathcal{H}_A to $\mathcal{H}_{BE} = \mathcal{H}_B \otimes \mathcal{H}_E$ and \mathcal{H}_E can be thought of as the “environment” of the channel. Transmission of the state S_A produces the state of the environment

$$S_E = \text{Tr}_B V S_A V^* = \tilde{\Phi}[S_A], \quad (7.45)$$

where $\tilde{\Phi}$ is the complementary channel from the input to the environment. In what follows we will simplify notations for the entropies by omitting the notation for the state, e.g. replace $H(S_A)$ by $H(A)$ etc., as we already did in the previous section.

Definition 7.28. The quantity $H(S_E) = H(\tilde{\Phi}[S_A])$, which is equal to the output entropy of the environment, is called the *entropy exchange* and will be denoted $H(S, \Phi)$, or simply $H(E)$.

By using the construction of the complementary channel from (6.42), we see that S_E is given by the density matrix $[\text{Tr } S V_j^* V_k]_{j,k=1,\overline{N}}$. Hence, we obtain the following result:

Proposition 7.29. *Let $\Phi[S] = \sum_{j=1}^N V_j S V_j^*$, with $\sum_{j=1}^N V_j^* V_j = I$. Then*

$$H(S, \Phi) = H\left([\text{Tr } S V_j^* V_k]_{j,k=1,\overline{N}}\right). \quad (7.46)$$

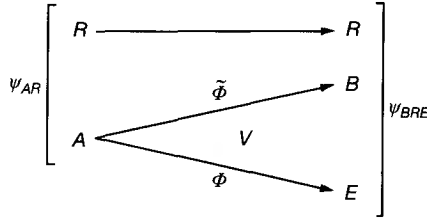


Figure 7.1. Purification with the reference system.

To obtain another useful expression for $H(S, \Phi)$, we take a reference system \mathcal{H}_R . Then we can purify the state S_A to $|\psi_{AR}\rangle\langle\psi_{AR}|$ in $\mathcal{H}_A \otimes \mathcal{H}_R$. Next, compute

$$H((\Phi \otimes \text{Id}_R)|\psi_{AR}\rangle\langle\psi_{AR}|) = H(S_{BR}). \quad (7.47)$$

Now, the tripartite system BRE described by the Hilbert space $\mathcal{H}_{BRE} = \mathcal{H}_B \otimes \mathcal{H}_R \otimes \mathcal{H}_E$ is in the pure state

$$|\psi_{BR}\rangle \otimes |\psi_E\rangle = (V \otimes I_R)|\psi_{AR}\rangle. \quad (7.48)$$

Looking at the split $BR|E$ we have a bipartite system in the pure state and therefore, by (7.30)

$$H(S_E) = H(S_{BR}). \quad (7.49)$$

Formula (7.49) provides an alternative expression for the entropy exchange as the output entropy of the extended channel $\Phi \otimes \text{Id}_R$ applied to the purified input state S_{AR} . Since the left-hand side does not involve R , this also shows that $H(S_{BR})$ does not depend on a particular way of purification of the state $S = S_A$.

The entropy exchange satisfies a simple but useful inequality: *Let $S = \sum_j p_j S_j$ be a mixture of pure states S_j . In this case,*

$$H(S, \Phi) \geq \sum_j p_j H(\Phi[S_j]). \quad (7.50)$$

Indeed, by concavity of the quantum entropy,

$$H(S, \Phi) = H\left(\tilde{\Phi}\left[\sum_j p_j S_j\right]\right) \geq \sum_j p_j H(\tilde{\Phi}[S_j]).$$

But $H(\tilde{\Phi}[S_j]) = H(\Phi[S_j])$, because the system BE is in a pure state as soon as the input S_j is pure.

Example 7.30. Consider the depolarizing channel (6.49) and the input chaotic state $\bar{S} = I_d/d$. By using representation (6.54) with the Kraus operators $V_{\alpha\beta}$ and the properties of discrete Weyl operators, we find

$$\text{Tr } \bar{S} V_{\alpha\beta}^* V_{\alpha'\beta'} = \delta_{(\alpha\beta),(\alpha'\beta')} \begin{cases} \frac{p}{d^2}, & (\alpha\beta) \neq (00), \\ 1 - p \frac{d^2-1}{d^2}, & (\alpha\beta) = (00). \end{cases}$$

Hence,

$$H(\bar{S}, \Phi) = - \left(1 - p \frac{d^2-1}{d^2}\right) \log \left(1 - p \frac{d^2-1}{d^2}\right) - p \frac{d^2-1}{d^2} \log \frac{p}{d^2}. \quad (7.51)$$

7.7 Quantum mutual information

So far, we have encountered three entropic quantities: the input entropy $H(S)$, the output entropy $H(\Phi[S])$ and the entropy exchange $H(S, \Phi)$. The relation between them allows us to interpret them as the sides of a triangle: the sum of any two of them is larger than the third. For example, the difference in length of the upper two sides of the triangle and its basis can be seen as the information correlation between R and B :

$$\begin{aligned} I(S, \Phi) &= H(S) + H(\Phi[S]) - H(S, \Phi) \\ &= H(A) + H(B) - H(E) \\ &= H(R) + H(B) - H(BR) \geq 0 \end{aligned} \quad (7.52)$$

by the subadditivity of the quantum entropy.

Since R has the same entropy as A , this quantity can be thought of as the *quantum mutual information* between the input and output. The mutual information $I(S, \Phi)$ has a number of nice properties, similar to that of the Shannon information.

Proposition 7.31. *The quantum mutual information $I(S, \Phi)$ is*

- i. *concave in S ;*
- ii. *convex in Φ ;*
- iii. *subadditive: $I(S_{12}, \Phi_1 \otimes \Phi_2) \leq I(S_1, \Phi_1) + I(S_2, \Phi_2)$;*
- iv. *satisfies the data processing inequalities:*

$$I(S, \Phi_2 \circ \Phi_1) \leq \min\{I(S, \Phi_1), I(\Phi_1(S), \Phi_2)\}.$$

Proof. The proof is based on playing with different splits of the tripartite system BRE , which is in the pure state (7.48), and on the properties of the conditional entropy.

$$\begin{aligned}
 \text{i.} \quad I(S, \Phi) &= H(R) + H(B) - H(E) \\
 &= H(BE) + H(B) - H(E) \\
 &= H(B|E) + H(B),
 \end{aligned} \tag{7.53}$$

where the first term is concave in S_{BE} by Corollary 7.27, and the second is concave in S_B . It remains for us to observe that the maps $S \rightarrow S_{BE}$ and $S \rightarrow S_B$ are affine.

$$\begin{aligned}
 \text{ii.} \quad I(S, \Phi) &= H(R) + H(B) - H(E) \\
 &= H(R) + H(B) - H(BR) \\
 &= H(R) - H(R|B),
 \end{aligned}$$

where the first term does not depend on Φ , while the second is concave in $S_{BR} = (\Phi \otimes \text{Id}_R)[S_{AR}]$, which is affine in Φ .

iii. This follows from the expression $I(S, \Phi) = H(B|E) + H(B)$, see (7.53), the subadditivity (7.43) of the conditional entropy $H(B|E)$ and the subadditivity (7.4) of the quantum entropy $H(B)$.

iv. Denote by A the input of Φ_1 , by B the common output of Φ_1 and input of Φ_2 , and by C the output of Φ_2 . Let also $E_{1,2}$ denote the environments for $\Phi_{1,2}$. We have

$$I(S, \Phi_2 \circ \Phi_1) = H(R) + H(C) - H(E_1 E_2). \tag{7.54}$$

To prove the first data processing inequality, note that

$$I(S, \Phi_1) = H(R) + H(B) - H(E_1), \tag{7.55}$$

so we have to show that $H(C) - H(E_1 E_2) \leq H(B) - H(E_1)$. But since the systems $CRE_1 E_2$ and BRE_1 are in pure states, this is equivalent to

$$H(R|E_1 E_2) \leq H(R|E_1),$$

which is true because of the monotonicity of the conditional entropy.

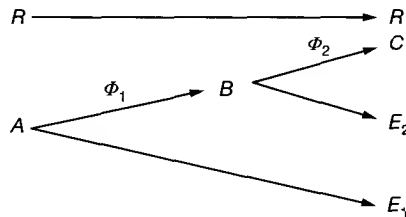


Figure 7.2. Composition of channels.

To prove the second data processing inequality, note that

$$I(\Phi_1[S], \Phi_2) = H(B) + H(C) - H(E_2) \quad (7.56)$$

so we have to show that $H(R) - H(E_1 E_2) \leq H(B) - H(E_2)$. But since the systems BRE_1 and $CRE_1 E_2$ are in pure states, this is the same as $H(R) - H(CR) \leq H(RE_1) - H(CRE_1)$ or

$$H(C|RE_1) \leq H(C|R).$$

However, the last inequality holds due to the monotonicity of the conditional entropy. \square

7.8 Notes and references

1. The entropy of a density operator was introduced by von Neumann [212]. A detailed survey of the properties of the quantum entropy and the quantum relative entropy can be found in the article of Wehrl [215] and in the books of Ohya and Petz [161] and Petz [168].

The inequality (7.3) can be strengthened by replacing $\text{Tr}(S - T)^2$ by $\|S - T\|_2^2$, see [180]. The proof is based on the corresponding classical Pinsker's inequality, see Lemma 12.6.1 in [42].

2. A real function g , defined on the interval $I \subseteq \mathbb{R}$, is called *operator-monotone* if, for arbitrary natural n and Hermitian $n \times n$ -matrices $A \leq B$ with spectra in I , the inequality $g(A) \leq g(B)$ holds. Operator-convex functions are defined in a similar fashion. The main results concerning operator-monotone and operator-convex functions, as well as a number of useful matrix inequalities can be found in the book of Bhatia [27]. By using the classical results from the theory of operator-monotone and operator-convex functions one can show that the class \mathcal{G} , defined by the relation (7.7), consists of operator-convex functions g on $(0, \infty)$ such that $g(1) = 0$.

There are several proofs of the Monotonicity Theorem 7.6. A usual approach is to derive it as a corollary of the celebrated ‘‘Lieb concavity’’ [161], [158]. In a series of papers [146], [147], and [148], Lindblad obtained a number of approximations to the monotonicity property and finally proved its equivalence to the strong subadditivity of the quantum entropy, established earlier by Lieb and Ruskai [145]. Uhlmann [209] gave a different proof based on an interpolation method. Quantum ‘‘quasi-entropies’’, including the relative g -entropies, were introduced by Petz [167], who gave a proof of Theorem 7.7 for maps satisfying (6.5), based on an operator generalization of the Jensen inequality. Here, we follow the paper of Lesniewski and Ruskai [143], where a direct proof of the monotonicity was given that makes no use of the Lieb concavity. This proof is based, after all, on a generalization of the Cauchy–Schwarz inequality and allows us to establish the monotonicity of a whole class of invariants of pairs of

states in quantum “geometro-statistics”, initiated by Morozova and Chentzov [39]. Effros pointed to the connection of these problems with the notion of the matrix perspective of an operator-convex function [56]. The condition for the equality in the monotonicity property was established by Petz (see also the survey of Ruskai [173]).

A new, “natural” proof of the monotonicity property of the relative entropy was suggested by Bjelaković and Siegmund-Schultze [28] (see also Hayashi [78]), based on the operational interpretation of the quantum relative entropy as the optimal error exponent in the problem of asymptotic discrimination between two quantum states (for the quantum analog of Stein’s Lemma in mathematical statistics, see Theorem 12.8.1 in [42]).

The entropy gain was introduced and studied by Alicki [6].

3. In this section we follow Lindblad [148], see also the book of Ohya and Petz [161].

4. Lemma 7.19 is due to Fannes [58]. Audenaert [12] obtained the sharp bound

$$|H(S_1) - H(S_2)| \leq \log(d-1) \cdot \|S_1 - S_2\|_1 + h_2(\|S_1 - S_2\|_1).$$

The connection of this inequality with Fano’s Lemma 4.14 is indicated in the book of Petz [168].

5. The information correlation was considered by Stratonovich [203] and by Lindblad [146]. The entanglement of formation was introduced in the fundamental work of Bennett, DiVincenzo, Smolin and Wootters [21]. For mixed states, there is a whole variety of measures of entanglement. The quantitative theory of entanglement is another separate chapter of quantum information science, in which the mathematical methods are substantially applied and developed, see e.g. the surveys of Keyl [128], Alber et al. [5], and the book of Hayashi [78].

The definition and properties of the convex hull, as well as the general duality theorem of convex programming are presented in, e.g. the books of Rockafellar [172], and Magaril–Il’yaev and Tikhomirov [154]. The proof of upper semicontinuity of the entanglement of formation is based on a construction from the work of Shirokov [185].

The quantum conditional entropy was introduced by Adami and Cerf [2]. The operational interpretation of the negative values of the conditional entropy as a “credit” for future quantum communication based on state merging protocol was proposed by M. Horodecki, Oppenheim, and Winter [120].

6. Entropy exchange was introduced (without using this name) by Lindblad [149] and later, independently, in the context of quantum information theory, by Barnum, Nielsen and Schumacher [16].

7. The quantum mutual information also first appeared in the paper of Lindblad [149] and later, independently, in the paper of Adami and Cerf [2], who studied its properties in detail in an information-theoretic context.

Part IV

Basic channel capacities

Chapter 8

The classical capacity of quantum channel

8.1 The coding theorem

In Chapter 5, the basic coding theorem for a classical-quantum communication channel was established. In this chapter, we will use this result to obtain the coding theorem for an arbitrary quantum channel.

Consider the channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ and the corresponding composite, “memoryless” channel $\Phi^{\otimes n} = \Phi \otimes \cdots \otimes \Phi : \mathfrak{T}(\mathcal{H}_A^{\otimes n}) \rightarrow \mathfrak{T}(\mathcal{H}_B^{\otimes n})$. The block code for such a composite channel comprises a c-q channel $i \rightarrow S_i^{(n)}$ encoding classical messages i into input states $S_i^{(n)}$ in the space $\mathcal{H}_A^{\otimes n}$, and a q-c channel (observable $M^{(n)}$) in the space $\mathcal{H}_B^{\otimes n}$, decoding the output states $\Phi^{\otimes n}[S_i^{(n)}]$ into classical messages j :

$$i \rightarrow S_i^{(n)} \rightarrow \Phi^{\otimes n}[S_i^{(n)}] \rightarrow j$$

This leads to the following definition

Definition 8.1. A code $(\Sigma^{(n)}, M^{(n)})$ of length n and of size N for the composite channel $\Phi^{\otimes n}$ consists of an *encoding*, given by a collection of states $\Sigma^{(n)} = \{S_i^{(n)}; i = 1, \dots, N\}$ in $\mathcal{H}_A^{\otimes n}$ and a *decoding*, described by an observable $M^{(n)} = \{M_j^{(n)}; j = 0, 1, \dots, N\}$ in $\mathcal{H}_B^{\otimes n}$.

The maximal error probability of the code is equal to

$$P_e(\Sigma^{(n)}, M^{(n)}) = \max_{i=1, \dots, N} [1 - p_{\Sigma M}(i|i)], \quad (8.1)$$

where

$$p_{\Sigma M}(j|i) = \text{Tr } \Phi^{\otimes n}[S_i^{(n)}] M_j^{(n)} \quad (8.2)$$

is the probability to make a decision j under the condition that a message i was sent. The minimum of the error $P_e(\Sigma^{(n)}, M^{(n)})$ over all codes of length n and size N is again denoted $p_e(n, N)$. The classical capacity $C(\Phi)$ of the quantum channel Φ is defined as the least upper bound of the rates R for which

$$\lim_{n \rightarrow \infty} p_e(n, 2^{nR}) = 0.$$

Let us recall that a finite probability distribution π on the set of quantum states $\mathfrak{S}(\mathcal{H})$, that assigns the probabilities π_i to states S_i , is called an *ensemble*. If an ensemble $\pi^{(n)}$, with probabilities $\{\pi_i^{(n)}\}$ of the input states $S_i^{(n)}$ is given, then using the transition probability $p_{\Sigma M}(j|i)$, we can find the joint distribution of input i and output j and compute the Shannon information $\mathcal{I}_n(\pi^{(n)}, M^{(n)})$ according to formula (5.14). Applying Shannon's Coding Theorem, we obtain, as in Proposition 5.16,

$$C(\Phi) = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{\{\pi^{(n)}, M^{(n)}\}} \mathcal{I}_n(\pi^{(n)}, M^{(n)}). \quad (8.3)$$

In contrast to the case of c-q channel, there is no fixed alphabet here, and one has to optimize not only with respect to the output observable $M^{(n)}$ and the input distribution $\{\pi_i^{(n)}\}$, but also with respect to all possible states $S_i^{(n)}$ at the input of the channel $\Phi^{\otimes n}$.

Proposition 8.2. *The classical capacity of channel Φ is equal to*

$$C(\Phi) = \lim_{n \rightarrow \infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}), \quad (8.4)$$

where

$$C_\chi(\Phi) = \sup_{\pi} \chi(\{\pi_i\}; \{\Phi[S_i]\}), \quad (8.5)$$

the quantity χ is defined by the relation (5.8) and the supremum is taken over all input ensembles $\pi = \{\pi_i, S_i\}$ in $\mathfrak{S}(\mathcal{H})$.

Note that, similar to Proposition 5.16, the limit in the relations (8.3), (8.4) as $n \rightarrow \infty$ is equal to the supremum over n , due to superadditivity of the corresponding sequences.

Proof. From the upper bound of Theorem 5.9

$$\mathcal{I}_n(\pi^{(n)}, M^{(n)}) \leq C_\chi(\Phi^{\otimes n}).$$

The inequality \leq in (8.4) then follows from (8.3).

Let us show that

$$C(\Phi) \geq \lim_{n \rightarrow \infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}) \equiv \bar{C}(\Phi). \quad (8.6)$$

Take $R < \bar{C}(\Phi)$. In this case, we can choose n_0 and an ensemble $\pi^{(n_0)} = \{\pi_i^{(n_0)}, S_i^{(n_0)}\}$ in $\mathcal{H}_A^{\otimes n_0}$ such that

$$n_0 R < \chi\left(\left\{\pi_i^{(n_0)}\right\}; \left\{\Phi^{\otimes n_0}[S_i^{(n_0)}]\right\}\right). \quad (8.7)$$

Consider the c-q channel $\tilde{\Phi}$ in $\mathcal{H}_B^{\otimes n_0}$ defined by the formula

$$i \rightarrow \Phi^{\otimes n_0}[S_i^{(n_0)}]. \quad (8.8)$$

By the Coding Theorem 5.4 for c-q channels, the capacity of $\tilde{\Phi}$ is

$$C(\tilde{\Phi}) = \max_{\{\pi_i\}} \chi \left(\{\pi_i\}; \left\{ \Phi^{\otimes n_0} \left[S_i^{(n_0)} \right] \right\} \right), \quad (8.9)$$

where the states $S_i^{(n_0)}$ are fixed and the maximum is taken over the probability distributions $\{\pi_i\}$. By (8.7), this is greater than $n_0 R$. Denoting by $\tilde{p}_e(n, N)$ the minimal error probability for $\tilde{\Phi}$, we have

$$p_e(n n_0, 2^{(n n_0) R}) \leq \tilde{p}_e(n, 2^{n(n_0 R)}), \quad (8.10)$$

since every code of size N for $\tilde{\Phi}$ is also a code of the same size for Φ . Thus, having chosen $R < \tilde{C}(\Phi)$, we can make the right hand side and hence the left hand side of (8.10) tend to zero as $n \rightarrow \infty$. Via the same argument as in Proposition 5.16 we can prove that $p_e(n', 2^{n' R}) \rightarrow 0$, when $n' \rightarrow \infty$, and hence (8.6) follows. \square

8.2 The χ -capacity

We shall call the quantity $C_\chi(\Phi)$, defined in (8.5), the χ -capacity of the channel Φ . Thus,

$$C_\chi(\Phi) = \sup_{\pi} \left[H \left(\sum_i \pi_i \Phi[S_i] \right) - \sum_i \pi_i H(\Phi[S_i]) \right], \quad (8.11)$$

where $\pi = \{\pi_i, S_i\}$. It can also be rewritten as

$$C_\chi(\Phi) = \sup_{S \in \mathcal{S}(\mathcal{H})} \left[H(\Phi(S)) - \hat{H}_\Phi(S) \right], \quad (8.12)$$

where the new channel characteristic

$$\hat{H}_\Phi[S] = \inf_{\pi: \sum_i \pi_i S_i = S} \sum_i \pi_i H(\Phi[S_i]) \quad (8.13)$$

was introduced, which is called the *convex hull* of the output entropy $H(\Phi[S])$. Here, the infimum is taken over all ensembles $\pi = \{\pi_i, S_i\}$, with the fixed average state $\sum_i \pi_i S_i = S$. Note that, by concavity of the quantum entropy, it is sufficient to take the infimum with respect to the convex decompositions of S into the *pure* states S_i .

Exercise 8.3. Prove that for the ideal channel $\hat{H}_{\text{Id}}(S) \equiv 0$. Hint: Consider the spectral decomposition of S .

The quantity (8.13) is closely related to the entanglement of formation (7.32), namely

$$\hat{H}_\Phi(S) = E_F(V S V^*),$$

where V is the isometry in the Stinespring representation (6.7) of the channel Φ , so that VSV^* is the state of the composite system “output-environment”. The proof of the following lemma is similar to that of Proposition 7.24, with the dimensionality $d_1 d_2$ of $\mathcal{H}_1 \otimes \mathcal{H}_2$ replaced with the dimensionality d_1 of \mathcal{H}_1 .

Lemma 8.4. *The convex hull of the output entropy $\hat{H}_\Phi(S)$ is a continuous convex function on $\mathfrak{S}(\mathcal{H})$. The infimum in (8.13) is attained on an ensemble consisting of no more than d_A^2 pure states, $d_A = \dim \mathcal{H}_A$.*

Corollary 8.5. *The supremum in the expression (8.11) for the χ -capacity is attained on an ensemble consisting of no more than d_A^2 pure states.*

Exercise 8.6. By using Exercise 4.8, show that an ensemble $\pi = \{\pi_i, S_i\}$ with the average state $\bar{S}_\pi = \sum_i \pi_i S_i$ is optimal for the supremum in expression (8.11) if and only if the following *maximal distance* condition holds: there exists a positive μ such that

$$H(\Phi[S]; \Phi[\bar{S}_\pi]) \leq \mu, \quad \text{for all (pure) input states } S, \quad (8.14)$$

with equality for members $S = S_i$ of the ensemble with $\pi_i > 0$. In this case, necessarily, $\mu = C_\chi(\Phi)$. Hint:

$$\frac{\partial}{\partial \pi_j} \chi(\{\pi_i\}; \{\Phi[S_i]\}) = H(\Phi[S_j]; \Phi[\bar{S}_\pi]) - \log e.$$

The χ -capacity can be computed for a number of interesting channels.

Exercise 8.7. Show that $C_\chi(\text{Id}) = \log d$ for the ideal channel in d -dimensional Hilbert space. For the quantum erasure channel $C_\chi(\Phi_p) = (1 - p) \log d$, where p is the probability of erasure. In both cases the optimal ensemble consists of equiprobable pure states corresponding to an orthonormal basis in \mathcal{H} .

Proposition 8.8. *Let Φ be a covariant channel (see Section 6.7),*

$$\Phi[U_g^A S (U_g^A)^*] = U_g^B \Phi[S] (U_g^B)^*.$$

If the representation U_g^A is irreducible (in which case Φ is called irreducibly covariant),

$$C_\chi(\Phi) = H\left(\Phi\left[\frac{I_A}{d_A}\right]\right) - \min_S H(\Phi[S]), \quad (8.15)$$

where the minimum is taken over the set of pure states.

Proof. Assume first that the symmetry group is finite. Irreducibility of the representation means that it has no nontrivial invariant subspaces, which is equivalent to the following: any operator that commutes with all U_g^A is a multiple of the unit operator. It follows, similar to the proof of relation (6.53), that for all states S

$$\frac{1}{|G|} \sum_{g \in G} U_g^A S (U_g^A)^* = \frac{I_A}{d_A}. \quad (8.16)$$

Let us show that

$$\max_{S \in \mathcal{E}(\mathcal{H})} H(\Phi[S]) = H\left(\Phi\left[\frac{I_A}{d_A}\right]\right). \quad (8.17)$$

This follows from the fact that the function $S \rightarrow H(\Phi[S])$ is concave and invariant with respect to the transformations $S \rightarrow U_g^A S U_g^{A*}$. Therefore, for arbitrary state S ,

$$\begin{aligned} H(\Phi[S]) &= |G|^{-1} \sum_{g \in G} H\left(\Phi\left[U_g^A S (U_g^A)^*\right]\right) \\ &\leq H\left(\Phi\left[|G|^{-1} \sum_{g \in G} U_g^A S (U_g^A)^*\right]\right). \end{aligned}$$

The inequality \leq in (8.15) then follows from (8.16), and it is sufficient to show that

$$C_X(\Phi) \geq H\left(\Phi\left[\frac{I_A}{d_A}\right]\right) - \min_S H(\Phi[S]). \quad (8.18)$$

Let us choose a state S_0 that minimizes the output entropy. Since the entropy is concave, it attains the minimum on pure states. Then the value in the right hand side of (8.15) is attained for the ensemble of states $S_g = U_g^A S_0 U_g^{A*}$; $g \in G$, with equal probabilities $\pi_g = |G|^{-1}$.

If the group is continuous, a similar argument applies, but the optimizing distribution will be continuous, namely the uniform distribution on G . One can then use a finite approximation and the continuity of the entropy. \square

Example 8.9. For the depolarizing channel (6.49)

$$C_X(\Phi) = \log d + \left(1 - p \frac{d-1}{d}\right) \log \left(1 - p \frac{d-1}{d}\right) + p \frac{d-1}{d} \log \frac{p}{d}, \quad (8.19)$$

which is achieved for an ensemble of equiprobable, orthogonal pure states forming an orthonormal basis in \mathcal{H} .

Indeed, the channel is irreducibly covariant, hence (8.15) applies. The output entropy is concave. Hence, it achieves the minimum on the set of pure states. For an

arbitrary pure input state, the output $\Phi[|\psi\rangle\langle\psi|]$ has the simple eigenvalue $(1 - p\frac{d-1}{d})$ and $d - 1$ eigenvalues $\frac{p}{d}$. Hence, the output entropy has the same value

$$-\left(1 - p\frac{d-1}{d}\right)\log\left(1 - p\frac{d-1}{d}\right) - p\frac{d-1}{d}\log\frac{p}{d}$$

for all pure input states, and (8.19) follows.

Example 8.10. Let us compute the χ -capacity of an arbitrary qubit unital channel Φ . Such a channel is irreducibly covariant with respect to the representation (6.65) (see Exercise 6.44). Hence, we can again apply Proposition 8.8.

In computing $\min_{S \in \mathfrak{S}(\mathcal{H})} H(\Phi(S))$ it is sufficient to consider $\Phi = \Lambda$, by taking into account the unitary invariance of the quantum entropy. We also remark that by Exercise 2.5 the entropy of the qubit state (2.8) is equal to

$$H(S(\vec{a})) = h_2\left(\frac{1 - |\vec{a}|}{2}\right). \quad (8.20)$$

Now, the Bloch ball is contracted, by the channel Λ , to the ellipsoid with lengths of the axes $|\lambda_\gamma|$, $\gamma = x, y, z$, and the minimal output entropy is attained at the top of the longest axis, which corresponds to the state with the eigenvalues $\frac{1 \pm \max_\gamma |\lambda_\gamma|}{2}$. This provides the χ -capacity

$$C_\chi(\Phi) = 1 - h_2\left(\frac{1 - \max_\gamma |\lambda_\gamma|}{2}\right). \quad (8.21)$$

8.3 The additivity problem

8.3.1 The effect of entanglement in encoding and decoding

Several interesting and difficult mathematical problems in quantum information theory are related to the proof or disproof of the *additivity property*:

$$C_\chi(\Phi_1 \otimes \Phi_2) = C_\chi(\Phi_1) + C_\chi(\Phi_2), \quad (8.22)$$

for some channels Φ_1, Φ_2 . Obviously, always

$$C_\chi(\Phi_1 \otimes \Phi_2) \geq C_\chi(\Phi_1) + C_\chi(\Phi_2).$$

If (8.22) is fulfilled for the channel $\Phi = \Phi_1$ and an arbitrary channel Φ_2 , then

$$C_\chi(\Phi^{\otimes n}) = nC_\chi(\Phi) \quad (8.23)$$

and due to relation (8.4) the classical capacity of the channel Φ is equal to the χ -capacity,

$$C(\Phi) = C_\chi(\Phi). \quad (8.24)$$

Property (8.22) was established for some important classes of channels $\Phi = \Phi_1$ (and arbitrary Φ_2) such as entanglement-breaking channels (Shor [190]), unital qubit channels (King [129]), depolarizing channels (King [130]), and several others. In particular, equality (8.24) holds for the classical capacity of all these channels. There was a strong belief that the additivity should hold for arbitrary channels Φ_1, Φ_2 , supported by the absence of counterexamples and intensive numerical search in low dimensions. However, it was finally shown that in very high dimensions there exist channels that violate (8.22), (8.23), (8.24). We will discuss these results later, in Section 8.3.5.

The additivity of the Shannon capacity for classical channels, after all, relies upon the absence of entanglement in the classical systems. To see this, let us try to generalize the classical proof based on the Kuhn–Tucker conditions (see Proposition 4.9) to the quantum case. Let Φ_1, Φ_2 be two channels with the optimal ensembles having averages $\bar{S}_{\pi_1}, \bar{S}_{\pi_2}$. We wish to prove that

$$C_X(\Phi_1 \otimes \Phi_2) \leq C_X(\Phi_1) + C_X(\Phi_2). \quad (8.25)$$

Then condition (8.14), applied to the product channel $\Phi_1 \otimes \Phi_2$, requires the following inequality

$$H((\Phi_1 \otimes \Phi_2)[S_{12}]; (\Phi_1 \otimes \Phi_2)[\bar{S}_{\pi_1} \otimes \bar{S}_{\pi_2}]) \leq C_X(\Phi_1) + C_X(\Phi_2) \quad (8.26)$$

to hold for *all* pure input states S_{12} of the product channel. Moreover, equality must hold for the tensor products of the members of the optimal ensembles with positive probabilities, which easily follows from the corresponding equalities for the channels Φ_1, Φ_2 , as well the inequality (8.26) for *product* states. However, proving (8.26) for *entangled* states is no simpler than proving the inequality

$$C_X(\Phi_1 \otimes \Phi_2) \leq C_X(\Phi_1) + C_X(\Phi_2),$$

which is the main problem with (8.22), since the opposite inequality follows from the definition of C_X .

Additivity of $C_X(\Phi)$ would have a surprising physical consequence – it would mean that using entangled input states does not increase the classical capacity of a quantum channel. However, as shown in Section 5.4, using entangled observables at the output of any essentially quantum channel (i.e. a channel with noncommuting output states) increases the capacity. From this point of view it is surprising that the additivity of $C_X(\Phi)$ is a rather widespread phenomenon. Let us discuss entangled inputs in more detail.

The encoding $\Sigma^{(n)} = \{S_i^{(n)}\}$ in the Definition 8.1 of a code will be called *unentangled* if

$$S_i^{(n)} = \sum_{x^n} p(x^n|i) S_{x_1} \otimes \cdots \otimes S_{x_n}, \quad (8.27)$$

where S_{x_k} is a state in the k -th copy of the space \mathcal{H}_A and $p(x^n|i)$ is a conditional probability describing preliminary processing of the input message. Let us recall that a decoding $M^{(n)}$ is unentangled if it has the form (5.37).

Given a quantum channel Φ , one can define the four classical capacities $C_{1,1}$, $C_{1,\infty}$, $C_{\infty,1}$, $C_{\infty,\infty}$, where the first (second) index refers to the inputs (outputs), ∞ means using arbitrary entangled inputs (outputs), and 1 means restriction to unentangled inputs (outputs) in Definition 8.1 and correspondingly in the formula (8.3). By definition,

$$C_{\infty,\infty} = C(\Phi).$$

Exercise 8.11. Show, by using (8.27), the data-processing inequality (4.20), and the Shannon Coding Theorem 4.13, that

$$C_{1,1} = \max_{\pi, M} \mathcal{I}_1(\pi, M).$$

Here, π is an arbitrary ensemble at the input of the channel Φ , and M is an arbitrary observable at the output.

The quantity $C_{1,1}$ is sometimes called the *Shannon capacity* of the quantum channel Φ .

Exercise 8.12. By using the Coding Theorem 5.4 for classical-quantum channels, show that

$$C_{1,\infty} = C_\chi(\Phi).$$

The relations between these four capacities are shown in the following diagram:

$$\begin{array}{ccc} C_{\infty,1} & \leq & C_{\infty,\infty} (= C) \\ \parallel & & \vee \\ C_{1,1} & \leq & C_{1,\infty} (= C_\chi) \end{array} \quad (8.28)$$

where \leq should be understood as “always less than or equal to, and strictly less for some channels”.

Equality

$$C_{1,1} = C_{\infty,1} \quad (8.29)$$

shows that using entangled input states with unentangled output measurements does not increase the accessible information. It will be proven below. The upper inequality $C_{\infty,1} \stackrel{<}{\neq} C_{\infty,\infty}$ follows from this and from the lower inequality, which expresses the strict superadditivity with respect to the output measurements demonstrated in Section 5.4.

Proof of the equality (8.29). Let Φ be a channel, let π be a probability distribution, assigning probability π_k to a state S_k at the input of the channel, and let $M = \{M_j\}$ be

the observable measured at the output of the channel. Denote by $\mathcal{I}_\Phi(\pi, M)$ the Shannon information corresponding to the input probability distribution $\{\pi_k\}$ and transition probability $p(j|k) = \text{Tr} \Phi[S_k] M_j = \text{Tr} S_k \Phi^*[M_j]$. In order to prove (8.29), it is sufficient to establish the inequality

$$\begin{aligned} \max_{\{\pi, M^1, M^2\}} \mathcal{I}_{\Phi_1 \otimes \Phi_2}(\pi, M^1 \otimes M^2) &\leq \max_{\{\pi^1, M^1\}} \mathcal{I}_{\Phi_1}(\pi^1, M^1) \\ &\quad + \max_{\{\pi^2, M^2\}} \mathcal{I}_{\Phi_2}(\pi^2, M^2), \end{aligned} \quad (8.30)$$

In terms of the classical mutual information

$$\mathcal{I}_\Phi(\pi, M) = I(X; Y),$$

where X is the input random variable, taking values k , and Y is the output random variable, taking values j . In order to prove (8.30), consider the states S_k in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the channel $\Phi_1 \otimes \Phi_2$, with a product observable $M^1 \otimes M^2$. In this case, the conditional probability is

$$p(j_1, j_2|k) = \text{Tr} S_k (\Phi_1^*[M_{j_1}^1] \otimes \Phi_2^*[M_{j_2}^2]) = p^1(j_1|j_2, k) p^2(j_2|k), \quad (8.31)$$

where

$$p^1(j_1|j_2, k) = \text{Tr} S_{j_2, k}^1 \Phi_1^*[M_{j_1}^1], \quad p^2(j_2|k) = \text{Tr} S_k^2 \Phi_2^*[M_{j_2}^2],$$

and

$$S_k^2 = \text{Tr}_1 S_k, \quad S_{j_2, k}^1 = \frac{\text{Tr}_2 S_k (I \otimes \Phi_2^*[M_{j_2}^2])}{\text{Tr} S_k (I \otimes \Phi_2^*[M_{j_2}^2])}.$$

Here, Tr_s denotes the partial trace in the s -th subsystem ($s = 1, 2$) in $\mathcal{H}_1 \otimes \mathcal{H}_2$.

We then have

$$I(X; Y_1 Y_2) = H(Y_1 Y_2) - H(Y_1 Y_2 | X),$$

where $H(\cdot)$, $H(\cdot|\cdot)$ are the entropy and conditional entropy, respectively, of the random variables describing the common classical input (X) and the outputs (Y_1, Y_2) of the two channels. By the subadditivity of the classical entropy,

$$H(Y_1 Y_2) \leq H(Y_1) + H(Y_2).$$

On the other hand,

$$H(Y_1 Y_2 | X) = H(Y_1 | Y_2 X) + H(Y_2 | X).$$

Combining, we get

$$I(X; Y_1 Y_2) \leq I(X Y_2; Y_1) + I(X; Y_2),$$

which, due to (8.31), amounts to

$$\mathcal{I}_{\Phi_1 \otimes \Phi_2}(\pi, M^1 \otimes M^2) \leq \mathcal{I}_{\Phi_1}(\pi^1, M^1) + \mathcal{I}_{\Phi_2}(\pi^2, M^2),$$

where π^1 is the probability distribution that assigns a probability $\pi_k p^2(j_2|k)$ to the state $S_{j_2, k}^1$, and π^2 is the probability distribution that assigns probability π_k to the state S_k^2 . Taking the maximum over π and S provides (8.30). \square

8.3.2 A hierarchy of additivity properties

An important characteristic of a quantum channel Φ , which already appeared in Section 8.2, is the minimal output entropy

$$\check{H}(\Phi) = \min_{S \in \mathfrak{S}(\mathcal{H})} H(\Phi[S]). \quad (8.32)$$

The corresponding hypothetical additivity property is

$$\check{H}(\Phi_1 \otimes \Phi_2) = \check{H}(\Phi_1) + \check{H}(\Phi_2). \quad (8.33)$$

By the concavity of the entropy, the minimum in (8.32) is attained on a pure state $S \in \text{extr}\mathfrak{S}(\mathcal{H})$. Therefore, the classical analog of the quantity $\check{H}(\Phi)$ is additive, because any pure state of a composite classical system is a product of the pure states of the subsystems. Instead, for a tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$ describing a composite quantum system, one has

$$\text{extr}\mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_2) \supsetneq \text{extr}\mathfrak{S}(\mathcal{H}_1) \times \text{extr}\mathfrak{S}(\mathcal{H}_2), \quad (8.34)$$

due to presence of the entangled states. Hence, there is no obvious reason for the additivity (8.33) in the quantum case.

Clearly, inequality \leq always holds in (8.33). This implies that (8.33) obviously holds for channels with zero minimal output entropy, i.e. those for which there exists a pure output state, such as ideal channel.

Exercise 8.13. Prove the following statement: the additivity properties (8.22) and (8.33) are equivalent for a pair of covariant channels satisfying the conditions of Proposition 8.8. Hint: use the fact that the tensor product of the irreducible representations of two symmetry groups G_1, G_2 is an irreducible representation of the group $G_1 \times G_2$.

Exercise 8.14. Consider the minimal entropy gain

$$\check{G}(\Phi) = \min_{S \in \mathfrak{S}(\mathcal{H})} (H(\Phi[S]) - H(S)).$$

Use the superadditivity property (7.21) to show that it is additive:

$$\check{G}(\Phi_1 \otimes \Phi_2) = \check{G}(\Phi_1) + \check{G}(\Phi_2). \quad (8.35)$$

Now consider the convex hull $\hat{H}_\Phi(S)$ of the output entropy, defined by relation (8.13). The hypothetical superadditivity property for this quantity is formulated in the following way: *for arbitrary state $S_{12} \in \mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ and channels Φ_1, Φ_2*

$$\hat{H}_{\Phi_1 \otimes \Phi_2}(S_{12}) \geq \hat{H}_{\Phi_1}(S_1) + \hat{H}_{\Phi_2}(S_2), \quad (8.36)$$

where S_1, S_2 are the partial traces of S_{12} in $\mathcal{H}_1, \mathcal{H}_2$.

This superadditivity property is *maximal* in the sense that its validity for a pair of channels Φ_1, Φ_2 implies all the other additivity properties for this pair of channels.

Proposition 8.15. *For two given channels Φ_1, Φ_2 the superadditivity property (8.36) implies both the additivity of the minimal output entropy (8.33) and of the χ -capacity (8.22).*

Proof. Indeed, let S_{12}^0 be a minimizer for $H((\Phi_1 \otimes \Phi_2)[S_{12}])$. In this case,

$$\begin{aligned} \check{H}(\Phi_1 \otimes \Phi_2) &= H((\Phi_1 \otimes \Phi_2)[S_{12}^0]) = \hat{H}_{\Phi_1 \otimes \Phi_2}(S_{12}^0) \\ &\geq \hat{H}_{\Phi_1}(S_1^0) + \hat{H}_{\Phi_2}(S_2^0) \geq \check{H}(\Phi_1) + \check{H}(\Phi_2), \end{aligned}$$

from which (8.33) follows. On the other hand, (8.36) and the subadditivity of the quantum entropy imply

$$\begin{aligned} &H((\Phi_1 \otimes \Phi_2)[S_{12}]) - \hat{H}_{\Phi_1 \otimes \Phi_2}(S_{12}) \\ &\leq H((\Phi_1 \otimes \Phi_2)[S_{12}]) - \hat{H}_{\Phi_1}(S_1) - \hat{H}_{\Phi_2}(S_2) \\ &\leq \left[H(\Phi_1[S_1]) - \hat{H}_{\Phi_1}(S_1) \right] + \left[H(\Phi_2[S_2]) - \hat{H}_{\Phi_2}(S_2) \right]. \end{aligned} \quad (8.37)$$

By using (8.12), we get

$$C_\chi(\Phi_1 \otimes \Phi_2) \leq C_\chi(\Phi_1) + C_\chi(\Phi_2),$$

i.e. (8.22). □

Moreover, the property (8.36) is closely related to the additivity of the χ -capacity with input constraints. Let F be a positive operator in the input Hilbert space \mathcal{H} of the channel, and E a positive constant. Consider the χ -capacity under the linear constraint $\text{Tr } SF \leq E$ on the input state S :

$$C_\chi(\Phi, F, E) = \max_{S: \text{Tr } SF \leq E} \left[H(\Phi[S]) - \hat{H}_\Phi(S) \right]. \quad (8.38)$$

Exercise 8.16. Assume that we have two channels Φ_1, Φ_2 with corresponding constraint operators F_1, F_2 . In this case, inequality (8.36), along with (8.12), implies

$$\begin{aligned} C_\chi(\Phi_1 \otimes \Phi_2, F_1 \otimes I_2 + I_1 \otimes F_2, E) &= \max_{E_1 + E_2 = E} \left[C_\chi(\Phi_1, F_1, E_1) \right. \\ &\quad \left. + C_\chi(\Phi_2, F_2, E_2) \right]. \end{aligned} \quad (8.39)$$

The property of superadditivity (8.36) is closely related to the following hypothetical property of the entanglement of formation: let $\mathcal{H}_1 = \mathcal{H}_1^A \otimes \mathcal{H}_1^B$ and $\mathcal{H}_2 = \mathcal{H}_2^A \otimes \mathcal{H}_2^B$, and let S_{12}^{AB} be a state in $\mathcal{H}_1 \otimes \mathcal{H}_2$, then

$$E_F(S_{12}^{AB}) \geq E_F(S_1^{AB}) + E_F(S_2^{AB}), \quad (8.40)$$

where the entanglement of formation is defined relative to the subdivision $A|B$. Namely, if property (8.40) would hold for all states S_{12}^{AB} , this would imply (8.36) for all channels Φ_1, Φ_2 and all states S_{12} , while if (8.36) would hold for all states S_{12} and partial trace channels, (8.40) would hold for all states S_{12}^{AB} .

The most complete conditional result in this direction is the following,

Theorem 8.17. *The conjectured properties (8.33), (8.22), (8.23) and (8.36) are globally equivalent in the following sense: assuming that one of them holds true for all channels Φ_1, Φ_2 , any other is also true for all channels. Moreover, they are globally equivalent to the superadditivity of the entanglement of formation (8.40) and to the additivity of the χ -capacity with arbitrary input constraints.*

When (parts of this) theorem were proved, there was hope that it would help to reduce the proof of more complicated global conjectures, such as additivity of the χ -capacity or superadditivity of entanglement of formation to a seemingly less complicated one – the global additivity of the minimal output entropy. However, with the disproof of the last conjecture, this theorem implies that none of the additivity properties holds globally (see Section 8.4 for more detail).

We will not consider the proof of this theorem here; instead we discuss some restricted classes of channels for which the maximal property (8.36) can be proved.

8.3.3 Some entropy inequalities

Here, we obtain inequalities that allow us to prove, in some cases, the superadditivity of the convex hull (8.36) and hence the additivity of the minimal output entropy (8.33) and of the χ -capacity (8.22).

Consider a state S and a non-ideal measurement, described by a collection of operators A_k that satisfy the normalization condition $\sum_k A_k^* A_k = I$, so that an outcome k occurs with probability $\pi_k = \text{Tr } S A_k^* A_k$, resulting in a posterior state of the system $S_k = A_k S A_k^* / \pi_k$. The entropies of the initial and posterior states are related by the Lindblad–Ozawa inequality

$$H(S) \geq \sum_k \pi_k H(S_k). \quad (8.41)$$

Proof. To prove this, take $\mathcal{H} = \mathcal{H}_1$ and consider a purification $|\psi\rangle$ of the state S in $\mathcal{H}_1 \otimes \mathcal{H}_2$. Then

$$\pi_k S_k = \text{Tr } \mathcal{H}_2(A_k \otimes I) |\psi\rangle \langle \psi| (A_k \otimes I)^*,$$

and $S_2 = \text{Tr } \mathcal{H}_1 |\psi\rangle\langle\psi| = \sum_k \pi_k S_2^k$, where

$$\pi_k S_2^k = \text{Tr } \mathcal{H}_1 (A_k \otimes I) |\psi\rangle\langle\psi| (A_k \otimes I)^*.$$

Using Theorem 3.10 twice and applying the concavity of the quantum entropy,

$$H(S) = H(S_2) = H\left(\sum_k \pi_k S_2^k\right) \geq \sum_k \pi_k H(S_2^k) = \sum_k \pi_k H(S_k). \quad \square$$

A particular case of the above is the following inequality for composite systems. Let S_{12} be a state of the composite system 12 and let $\{|e_k^2\rangle\}$ be an orthonormal basis in \mathcal{H}_2 . Taking $A_k = I_1 \otimes |e_k^2\rangle\langle e_k^2|$, one obtains

$$H(S_{12}) \geq \sum_k \pi_k H(S_1^k), \quad (8.42)$$

where $\pi_k S_1^k = \langle e_k^2 | S_{12} | e_k^2 \rangle$. This can be used to prove a particular case of the additivity conjecture.

Proposition 8.18. *The maximal property (8.36) holds in the case where $\Phi_1 = \Phi$ is an arbitrary channel in \mathcal{H}_1 and $\Phi_2 = \text{Id}_2$ is the ideal channel in \mathcal{H}_2 .*

Proof. Since $\hat{H}_{\text{Id}}(S) \equiv 0$, according to Exercise 8.3, we have to prove that

$$\hat{H}_{(\Phi \otimes \text{Id}_2)}(S_{12}) \geq \hat{H}_{\Phi}(S_1). \quad (8.43)$$

Let $S_{12} = \sum_i \pi_i S_{12}^i$ be the optimal decomposition, i.e.

$$\hat{H}_{(\Phi \otimes \text{Id}_2)}(S_{12}) = \sum_i \pi_i H((\Phi \otimes \text{Id}_2)[S_{12}^i]).$$

By using (8.42), we obtain

$$\sum_i \pi_i H((\Phi \otimes \text{Id}_2)[S_{12}^i]) \geq \sum_{ik} \pi_i \pi_{ik} H(\Phi[S_1^{ik}]), \quad (8.44)$$

where $\pi_{ik} S_1^{ik} = \langle e_k^2 | S_{12}^i | e_k^2 \rangle$. Now, we have the decomposition of S_1 into the states S_1^{ik} with probabilities $\pi_i \pi_{ik}$ for channel Φ . Hence, the right hand side of (8.44) is greater than or equal to $\hat{H}_{\Phi}(S_1)$. \square

Proposition 8.19. *The maximal property (8.36) (and hence the additivity of the minimal output entropy (8.33) and of the χ -capacity (8.22)) hold in the case, where $\Phi_1 = \Phi$ is an entanglement-breaking channel (see Section 6.4) and $\Phi_2 = \Psi$ is an arbitrary channel.*

Lemma 8.20. Let $\pi_j \geq 0$, $\sum \pi_j = 1$, and S_1^j, S_2^j be arbitrary states of the systems 1, 2. Then

$$H\left(\sum_j \pi_j S_1^j \otimes S_2^j\right) \geq H\left(\sum_j \pi_j S_1^j\right) + \sum_j \pi_j H(S_2^j).$$

Proof. We introduce the notation

$$(\bar{S}_{12})_\pi = \sum_j \pi_j S_1^j \otimes S_2^j; \quad (\bar{S}_1)_\pi = \sum_j \pi_j S_1^j.$$

Now, using the additivity of the entropy, the inequality can be rewritten as

$$H((\bar{S}_{12})_\pi) - \sum_j \pi_j H(S_1^j \otimes S_2^j) \geq H((\bar{S}_1)_\pi) - \sum_j \pi_j H(S_1^j),$$

which by identity (7.19) is equivalent to

$$\sum_j \pi_j H(S_1^j \otimes S_2^j; (\bar{S}_{12})_\pi) \geq \sum_j \pi_j H(S_1^j; (\bar{S}_1)_\pi).$$

However, the last inequality follows from the monotonicity of the relative entropy (with respect to partial trace in the second system). \square

Proof. Coming back to the proof of the proposition, assume that Φ is an entanglement-breaking channel, so that $\Phi[S] = \sum_j S_1^j \text{Tr} S M_1^j$, where $\{M_1^j\}$ is an observable in system 1. In this case,

$$(\Phi \otimes \text{Id}_2)[S_{12}] = \sum_j p_j S_1^j \otimes S_2^j, \quad (8.45)$$

where $p_j S_2^j = \text{Tr}_{\mathcal{H}_1} S_{12} (M_1^j \otimes I_2)$ for an arbitrary state S_{12} of the composite system. Taking the partial trace, we obtain, in particular,

$$\Phi[S_1] = \sum_j p_j S_1^j, \quad S_2 = \sum_j p_j S_2^j. \quad (8.46)$$

If Ψ is an arbitrary channel in system 2,

$$(\Phi \otimes \Psi)[S_{12}] = \sum_j p_j S_1^j \otimes \Psi[S_2^j]. \quad (8.47)$$

We have

$$\hat{H}_{\Phi \otimes \Psi}(S_{12}) = \inf_i \sum_i \pi_i H((\Phi \otimes \Psi)[S_{12}^i]), \quad (8.48)$$

where $S_{12} = \sum_i \pi_i S_{12}^i$ is an arbitrary decomposition. By writing relation (8.45) for S_{12}^i we obtain

$$(\Phi \otimes \text{Id}_2)[S_{12}^i] = \sum_j p_{ij} S_1^{ij} \otimes S_2^{ij},$$

along with the decompositions

$$S_1 = \sum_i \pi_i S_1^i, \quad S_2 = \sum_{ij} \pi_i p_{ij} S_2^{ij}. \quad (8.49)$$

By using (8.47), the lemma that we just proved, and (8.46), we obtain

$$\begin{aligned} H((\Phi \otimes \Psi)[S_{12}^i]) &= H\left(\sum_j p_{ij} S_1^{ij} \otimes \Psi[S_2^{ij}]\right) \\ &\geq H\left(\sum_j p_{ij} S_1^{ij}\right) + \sum_j p_{ij} H(\Psi[S_2^{ij}]) \\ &\geq H(\Phi[S_1^i]) + \sum_j p_{ij} H(\Psi[S_2^{ij}]), \end{aligned}$$

which in combination with (8.48) and (8.49) implies

$$\hat{H}_{\Phi \otimes \Psi}(S_{12}) \geq \hat{H}_{\Phi}(S_1) + \hat{H}_{\Psi}(S_2). \quad \square$$

8.3.4 Additivity for complementary channels

Lemma 8.21. *If S is a pure state, then it holds for any channel Φ ,*

$$H(\Phi[S]) = H(\tilde{\Phi}[S]), \quad (8.50)$$

where $\tilde{\Phi}$ is the complementary of Φ . Therefore, if either of the properties (8.36), (8.33) holds for channels Φ_1, Φ_2 , the similar property holds for the complementary channels $\tilde{\Phi}_1, \tilde{\Phi}_2$.

Proof. As follows from the definition (6.38) of the complementary channels, the states $\Phi[S]$ and $\tilde{\Phi}[S]$ are partial states of the pure state VSV^* . Hence, by equality (7.30), they have equal entropies.

From (8.50) we conclude

$$\check{H}(\Phi) = \check{H}(\tilde{\Phi}); \quad \hat{H}_{\Phi}(S) = \hat{H}_{\tilde{\Phi}}(S)$$

for an arbitrary state S , due to the expressions for $\check{H}(\Phi)$, $\hat{H}_{\Phi}(S)$ in terms of the output entropies for pure input states. Hence, the second statement follows. \square

Proposition 8.19 then implies

Corollary 8.22. *The maximal property (8.36) (and hence the additivity of the minimal output entropy (8.33) and of the χ -capacity (8.22)) hold in the case where $\Phi_1 = \Phi$ is the channel (6.45), in particular, a dephasing channel, and $\Phi_2 = \Psi$ is an arbitrary channel.*

Proposition 8.23. *Let Φ_2 be an arbitrary channel. The maximal property (8.36) holds for Φ_1, Φ_2 if Φ_1 is an orthogonal convex sum (see Definition 6.33) of the ideal channel and a channel $\Phi^{(0)}$ such that the maximal property holds for $\Phi^{(0)}$ and Φ_2 .*

It follows that such a Φ_1 fulfills the additivity of the minimal output entropy (8.33) and of the χ -capacity (8.22), with an arbitrary Φ_2 . For example, this holds for the erasure channel, because it is an orthogonal convex sum of the ideal channel and an entanglement-breaking channel for which the property (8.36) was established in Section 8.3.3.

Proof. Denote $\Phi^{(q)} = q\text{Id} \oplus (1-q)\Phi^{(0)}$. Now,

$$H(\Phi^{(q)}[S]) = qH(S) + (1-q)H(\Phi^{(0)}[S]) + h_2(q),$$

where $h_2(q) = -q \log q - (1-q) \log(1-q)$ is the binary entropy and

$$\hat{H}_{\Phi^{(q)}}(S) = (1-q)\hat{H}_{\Phi^{(0)}}(S) + h_2(q), \quad (8.51)$$

because the minimum in the expression for $\hat{H}_{\Phi^{(0)}}(S)$ is attained for an ensemble of pure states S_j for which $H(S_j) = 0$. For an arbitrary channel Φ_2 we have

$$\Phi^{(q)} \otimes \Phi_2 = q(\text{Id} \otimes \Phi_2) \oplus (1-q)(\Phi^{(0)} \otimes \Phi_2).$$

Then

$$\begin{aligned} \hat{H}_{\Phi^{(q)} \otimes \Phi_2}(S_{12}) &\geq q\hat{H}_{\text{Id} \otimes \Phi_2}(S_{12}) + (1-q)\hat{H}_{\Phi^{(0)} \otimes \Phi_2}(S_{12}) + h_2(q) \\ &\geq q\hat{H}_{\Phi_2}(S_2) + (1-q)[\hat{H}_{\Phi^{(0)}}(S_1) + \hat{H}_{\Phi_2}(S_2)] + h_2(q), \end{aligned}$$

where we have used the superadditivity of $\hat{H}_{\text{Id} \otimes \Phi_2}(S_{12})$, $\hat{H}_{\Phi^{(0)} \otimes \Phi_2}(S_{12})$ and the fact that $\hat{H}_{\text{Id}}(S_1) \equiv 0$. By using (8.51), we obtain that the right hand side is equal to $\hat{H}_{\Phi^{(q)}}(S_1) + \hat{H}_{\Phi_2}(S_2)$. \square

Corollary 8.22 implies that Proposition 8.23 holds with the replacement of the ideal channel by its complementary, the completely depolarizing channel.

8.3.5 Nonadditivity of quantum entropy quantities

The *quantum Rényi entropy* of order $p > 1$ of a density operator S is defined as

$$R_p(S) = \frac{1}{1-p} \log \text{Tr } S^p, \quad (8.52)$$

and in the limit $p \rightarrow 1$ the quantum Rényi entropies converge uniformly to the von Neumann entropy of a density operator S

$$\lim_{p \rightarrow 1} R_p(S) = -\text{Tr } S \log S \equiv H(S).$$

Defining the minimal output Rényi entropy of the channel Φ as

$$\check{R}_p(\Phi) = \min_{S \in \mathcal{G}(\mathcal{H})} R_p(\Phi[S]),$$

one may ask for an additivity property similar to (8.33), namely

$$\check{R}_p(\Phi_1 \otimes \Phi_2) = \check{R}_p(\Phi_1) + \check{R}_p(\Phi_2). \quad (8.53)$$

Again, inequality \leq is obvious here. Note that the validity of (8.53) for some specific channels Φ_1, Φ_2 and p close to 1 implies (8.33) for these Φ_1, Φ_2 . This raised interest in the problem (8.53) in the hope that its solution could shed light on the additivity problems in quantum information theory. In particular, proving it globally would imply, via Theorem 8.17, a global positive solution for all the additivity conjectures.

However, a counterexample of the *transpose-depolarizing channel* [216] soon appeared:

$$\Phi[S] = \frac{1}{d-1} (I - S^T) \quad (8.54)$$

for which (8.53) with $\Phi_1 = \Phi_2 = \Phi$ fails to hold for $d = \dim \mathcal{H} \geq 3$ and large enough p .

Exercise 8.24. Prove the complete positivity of the map (8.54) by establishing the Kraus representation

$$\Phi[S] = \frac{1}{2(d-1)} \sum_{j,k=1}^d (|e_j\rangle\langle e_k| - |e_k\rangle\langle e_j|) S (|e_k\rangle\langle e_j| - |e_j\rangle\langle e_k|), \quad (8.55)$$

where $\{e_j\}$ is an orthonormal basis.

A breakthrough in the negative solution of conjecture (8.53) came in 2007. Winter [225] had shown that in very high dimensions there always exist channels for which (8.53) does not hold for any $p > 2$. Next, Hayden [80] proved this for any

$1 < p < 2$. Winter shows that the additivity of the Rényi entropy with $p > 2$ breaks for the pair of channels $\Phi, \bar{\Phi}$ (complex conjugate in a fixed basis), where

$$\Phi[S] = \frac{1}{n} \sum_{j=1}^n U_j S U_j^*, \quad (8.56)$$

is the uniform mixture of unitary evolutions in \mathcal{H} and $n, d = \dim \mathcal{H}$ are large enough.

The proof relies upon the two estimates. A simple but efficient upper bound for $\check{R}_p(\Phi \otimes \bar{\Phi})$ is given by the inequality

$$\check{R}_p(\Phi \otimes \bar{\Phi}) \leq R_p(\Phi \otimes \bar{\Phi})[|\Omega\rangle\langle\Omega|] \leq \frac{p}{p-1} \log n, \quad (8.57)$$

where $|\Omega\rangle$ is the maximally entangled vector in the same basis. Indeed,

$$\begin{aligned} (\Phi \otimes \bar{\Phi})[|\Omega\rangle\langle\Omega|] &= \frac{1}{n^2} \sum_{j,k=1}^n (U_j \otimes \bar{U}_k) |\Omega\rangle\langle\Omega| (U_j \otimes \bar{U}_k)^* \\ &= \frac{1}{n} |\Omega\rangle\langle\Omega| + \frac{1}{n^2} \sum_{j \neq k} (U_j \otimes \bar{U}_k) |\Omega\rangle\langle\Omega| (U_j \otimes \bar{U}_k)^* \\ &\geq \frac{1}{n} |\Omega\rangle\langle\Omega|, \end{aligned}$$

where in the second equality we used the property $(U \otimes \bar{U})|\Omega\rangle = |\Omega\rangle$ for arbitrary unitary U . It follows that the maximal eigenvalue of the density operator $(\Phi \otimes \bar{\Phi})[|\Omega\rangle\langle\Omega|]$, i.e. its operator norm, is greater than or equal to $\frac{1}{n}$

$$\|(\Phi \otimes \bar{\Phi})[|\Omega\rangle\langle\Omega|]\| \geq \frac{1}{n}. \quad (8.58)$$

Hence,

$$R_p((\Phi \otimes \bar{\Phi})[|\Omega\rangle\langle\Omega|]) \leq \frac{1}{1-p} \log \left(\frac{1}{n} \right)^p = \frac{p}{p-1} \log n, \quad (8.59)$$

and (8.57) follows.

The next step is to consider random independent unitary operators $U_j; j = 1, \dots, n$, distributed according to the normalized Haar measure on the group of unitaries in \mathcal{H} . In this case, the expectation of each term in (8.56) is $EU_j S U_j^* = I/d$, since it is the only density operator that is invariant under unitary conjugations and one can expect that $\Phi[S]$ is close to I/d in a probabilistic sense for n large enough. A key ingredient of the proof is the large deviation estimate for the sums of random operators, inspired by the classical Bernstein–Chernoff–Hoeffding inequality, which implies that the probability of the inequality

$$\left\| \Phi[S] - \frac{I}{d} \right\| \leq \frac{\varepsilon}{d} \quad (8.60)$$

tends to 1 when $d \rightarrow \infty$, $n \sim d \log d$ (compare this result with Theorem 10.34). From inequality (8.60) it follows that $\|\Phi[S]\| \leq \frac{1+\varepsilon}{d}$ and hence

$$\mathrm{Tr} \Phi[S]^p = \mathrm{Tr} \Phi[S] \Phi[S]^{p-1} \leq \|\Phi[S]\|^{p-1} \mathrm{Tr} \Phi[S] \leq \left(\frac{1+\varepsilon}{d} \right)^{p-1} \quad (8.61)$$

so that

$$\check{R}_p(\Phi) = -\frac{1}{p-1} \max_{S \in \mathcal{S}(\mathcal{H})} \log \mathrm{Tr} \Phi[S]^p \geq \log \frac{d}{1+\varepsilon}.$$

This implies the probabilistic lower bound

$$\lim_{d \rightarrow \infty} \mathbb{P} \left\{ \check{R}_p(\Phi) \geq \log \frac{d}{1+\varepsilon} \right\} = 1$$

for random Φ of the form (8.56), with $n \sim d \log d$. Taking into account that $\check{R}_p(\Phi) = \check{R}_p(\bar{\Phi})$, and comparing with (8.57), one sees that with probability close to 1,

$$\check{R}_p(\Phi \otimes \bar{\Phi}) \leq \frac{p}{p-1} \log n < 2 \log \left(\frac{d}{1+\varepsilon} \right) \leq \check{R}_p(\Phi) + \check{R}_p(\bar{\Phi})$$

if $p > 2$, $d \rightarrow \infty$, $n \sim d \log d$.

The estimate (8.58) can be generalized to an arbitrary channel Φ as follows:

Lemma 8.25. *Let Φ be a quantum channel from A to B which has the Kraus representation with d_E components, and let $|\Omega_{AA}\rangle$ be the maximally entangled vector in $\mathcal{H}_A \otimes \mathcal{H}_A$. Then*

$$\|(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]\| \geq \frac{d_A}{d_B d_E} \equiv D. \quad (8.62)$$

Proof. We have

$$(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|] = \sum_{r,s=1}^{d_E} (V_r \otimes \bar{V}_s) |\Omega_{AA}\rangle\langle\Omega_{AA}| (V_r \otimes \bar{V}_s)^*$$

and

$$\langle\Omega_{BB}|(X \otimes Y)|\Omega_{AA}\rangle = \frac{1}{d_A d_B} \mathrm{Tr} Y^\top X,$$

for any operators X, Y acting from \mathcal{H}_A to \mathcal{H}_B .

By using these identities, we have

$$\begin{aligned}
 \|(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]\| &\geq \langle\Omega_{BB}|(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]|\Omega_{BB}\rangle \\
 &= \sum_{r,s=1}^{d_E} |\langle\Omega_{BB}|(V_r \otimes \bar{V}_s)|\Omega_{AA}\rangle|^2 \\
 &= \frac{1}{d_A d_B} \sum_{r,s=1}^{d_E} |\text{Tr } V_s^* V_r|^2 \\
 &\geq \frac{1}{d_A d_B} \sum_{r=1}^{d_E} |\text{Tr } V_r^* V_r|^2 \\
 &\geq \frac{1}{d_A d_B d_E} \left| \text{Tr} \sum_{r=1}^{d_E} V_r^* V_r \right|^2 \\
 &= \frac{1}{d_A d_B d_E} |\text{Tr } I_A|^2 = \frac{d_A}{d_B d_E},
 \end{aligned}$$

where the Cauchy–Schwarz inequality was used. \square

This estimate is one part of proving that in very high dimensions there exist channels that violate the additivity of the minimal Rényi entropies for all $p > 1$ (Winter and Hayden [83]). In this case, one considers the pair of channels $\Phi, \bar{\Phi}$, where Φ is a generic channel given by a general open system representation (6.22) with the random unitary evolution operator uniformly distributed over the group of unitaries. Moreover, $d_B = d_E = d \rightarrow \infty$, and $d_A \sim d^{1+1/p}$ so that $D \sim d^{1/p-1}$. The upper bound

$$\check{R}_p(\Phi \otimes \bar{\Phi}) \lesssim \log d \quad (8.63)$$

then follows from (8.62) similarly to (8.59).

The most difficult part of the proof is the probabilistic lower estimate

$$\lim_{d \rightarrow \infty} \mathbb{P} \left\{ \check{R}_p(\Phi) \geq \log d - c \right\} = 1, \quad (8.64)$$

which is also based on a large deviation phenomenon, namely the measure concentration, used to establish the concentration of the output entropy. This is closely related to the early observations that, given a random uniformly distributed pure state vector $|\psi_{AB}\rangle$ of a composite system AB , the entropy of the partial state S_B of the smaller system tends to its maximal value $\log d_B$ as $d_A \rightarrow \infty$, and hence S_B becomes almost chaotic while $|\psi_{AB}\rangle$ is almost maximally entangled.

The relations (8.63) and (8.64), together with $\check{R}_p(\Phi) = \check{R}_p(\bar{\Phi})$ imply that

$$\check{R}_p(\Phi \otimes \bar{\Phi}) < \check{R}_p(\Phi) + \check{R}_p(\bar{\Phi})$$

with probability close to 1.

Notably, these estimates for $p > 1$ are not strong enough to imply the violation of additivity for $p \rightarrow 1$ and the case of the minimal von Neumann entropy requires additional efforts which were accomplished by Hastings [76]. His original argument was improved and simplified in subsequent works [64], [30], and [11]. Following [30], consider the random unitary evolution channels with $d_B = d \rightarrow \infty$, and $d_E = d_A \gg d$. In this case $D = 1/d$, and Lemma 8.25 implies that the density operator $(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]$, acting in the space of dimensionality d^2 , has an eigenvalue greater than or equal to $1/d$.

Exercise 8.26. Let $P = [\lambda_1, \dots, \lambda_{d^2}]$ run through all probability distributions, such that $\lambda_1 \geq 1/d$. In this case,

$$\max_P H(P) = \log d + \frac{d-1}{d} \log(d+1) \leq 2 \log d - \frac{\log(d/e)}{d}$$

where the maximum is attained for the distribution P such that $\lambda_1 = 1/d$ and all other probabilities are equal to $1/(d+1)$.

It follows that

$$\check{H}(\Phi \otimes \bar{\Phi}) \leq H((\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]) \leq 2 \log d - \frac{\log(d/e)}{d} \quad (8.65)$$

which gives the “easier” part of the argument. The difficult part is again the probabilistic estimate, exploring at full strength the phenomenon of entropy concentration,

$$P \left\{ \check{H}(\Phi) > \log d - \frac{C}{d} \right\} > 0$$

for $d, d_A/d$ large enough, where C is a positive constant. Together with (8.65), this proves the existence of channels violating the additivity of the minimal von Neumann entropy.

Although, when combined with Theorem 8.17, this gives a definite negative answer to all global additivity conjectures, including the one which concerns the χ -capacity, several important issues remain open. All the above proofs use the technique of random unitary operators or random states and, as such, are not constructive. They only provide evidence for existence of counterexamples but do not allow us to actually produce them. Attempts at providing estimates for the dimensions in which nonadditivity can happen has so far led to overwhelmingly high values. The detailed estimates made in [64] gave $d \approx 3.9 \times 10^4$, $d_A \approx 7.8 \times 10^{32}$, breaking the additivity by a quantity

of the order of 10^{-5} . While this does not exclude the possibility of better estimates, perhaps to be based on a different (but yet unknown) approach, it casts doubt on finding concrete counterexamples by computer simulation of random unitary channels. It remains a mystery what happens in realistic dimensions. Perhaps, for some unknown reason, the additivity still holds generically or its violation is so tiny that it cannot be caught by numerical simulations.

From this point of view, the following explicit construction is of interest. Following [73], we shall consider an explicit example of a simple channel closely related to (8.54), for which the minimal Rényi entropies are nonadditive for all $p > 2$ and sufficiently large d . This more recent explicit construction thus achieves the same goal as the first Winter's proposal (8.56). However, it is not clear if it could be extended to the most interesting range $p \geq 1$.

Exercise 8.27. By using formula (6.42) and the Kraus representation (8.55), show that the complementary channel of (8.54) is

$$\bar{\Phi}[S] = \frac{2}{(d-1)} P_{-}(S \otimes I_2) P_{-}, \quad (8.66)$$

where $P_{-} = \frac{I_{12} - F}{2}$ is the projector onto the antisymmetric subspace \mathcal{H}_{-} of $\mathcal{H} \otimes \mathcal{H}$, which has dimensionality $\frac{d(d-1)}{2}$, and F is the flip operator acting as

$$F(|\psi_1\rangle \otimes |\psi_2\rangle) = |\psi_2\rangle \otimes |\psi_1\rangle \quad (8.67)$$

By Lemma 8.21, this channel shares the additivity properties with the channel (8.54).

Now consider the completely positive map $\frac{(d-1)}{2} \bar{\Phi}^*$, where $\bar{\Phi}^*$ is the dual to the channel (8.66):

$$S_{12} \longrightarrow \frac{(d-1)}{2} \bar{\Phi}^*[S_{12}] = \text{Tr}_2 P_{-} S_{12} P_{-}. \quad (8.68)$$

Its restriction to operators with support in the subspace \mathcal{H}_{-} is trace-preserving and hence it is a channel, which we denote Φ_{-} . For this channel, the input dimensionality is $\dim \mathcal{H}_{-} = \frac{d(d-1)}{2}$, while the output dimensionality is d and the environment dimensionality also is d , because (8.68) involves the partial trace with respect to the second copy of \mathcal{H} . Thus, we have $d_A = \frac{d(d-1)}{2}$, $d_B = d_E = d$, whence $D = \frac{(d-1)}{2d}$ in (8.62), so that

$$\text{Tr}(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]^p \geq \|(\Phi \otimes \bar{\Phi})[|\Omega_{AA}\rangle\langle\Omega_{AA}|]\|^p \geq D^p,$$

and

$$\check{R}_p(\Phi \otimes \bar{\Phi}) \leq \frac{p}{p-1} \log D. \quad (8.69)$$

The channel (8.68) breaks the additivity of the quantum Rényi entropies for $p > 2$:

Proposition 8.28. *For any $p > 2$ and $d > \left(1 - 2^{2/p-1}\right)^{-1}$*

$$\check{R}_p(\Phi_- \otimes \Phi_-) < \check{R}_p(\Phi_-) + \check{R}_p(\bar{\Phi}_-).$$

Proof. Let us show that

$$\check{R}_p(\Phi_-) = 1. \quad (8.70)$$

Let $S_{12} = |\psi_{12}\rangle\langle\psi_{12}|$ be a pure input state of the channel (8.68), where the vector $|\psi_{12}\rangle \in \mathcal{H}_-$ has the Schmidt decomposition $|\psi_{12}\rangle = \sum_{j=1}^d \sqrt{\lambda_j} |e_j \otimes h_j\rangle$. From the definition (8.68) of the channel,

$$\Phi_-[S_{12}] = S_1 = \sum_{j=1}^d \lambda_j |e_j\rangle\langle e_j|.$$

We have

$$\begin{aligned} \lambda_j &= |\langle\psi_{12}|e_j \otimes h_j\rangle|^2 \\ &\leq \langle e_j \otimes h_j | P_- | e_j \otimes h_j \rangle \\ &= \frac{1}{2} \langle e_j \otimes h_j | (I_{12} - F) | e_j \otimes h_j \rangle \\ &= \frac{1}{2} (1 - |\langle e_j | h_j \rangle|^2) \leq \frac{1}{2}, \end{aligned}$$

with the equality in the last inequality if $\langle e_j | h_j \rangle = 0$. Therefore,

$$\|\Phi_-[S_{12}]\| = \max_j \lambda_j \leq \frac{1}{2},$$

Hence, similar to (8.61),

$$\mathrm{Tr} \Phi_-[S_{12}]^p = \mathrm{Tr} \Phi_-[S_{12}] \Phi_-[S_{12}]^{p-1} \leq \|\Phi_-[S_{12}]\|^{p-1} \mathrm{Tr} \Phi_-[S_{12}] \leq \left(\frac{1}{2}\right)^{p-1}$$

and $R_p(\Phi_-[S_{12}]) = \frac{1}{1-p} \log \mathrm{Tr} \Phi_-[S_{12}]^p \geq 1$. The equality is attained for $|\psi_{12}\rangle = \frac{1}{\sqrt{2}}(|e\rangle \otimes |h\rangle - |h\rangle \otimes |e\rangle)$, where $\langle e | h \rangle = 0$. This proves (8.70).

Now, (8.69) together with $\Phi = \bar{\Phi}$ implies

$$\check{R}_p(\Phi_- \otimes \Phi_-) \leq \frac{p}{p-1} \log \frac{(d-1)}{2d} < 2 = \check{R}_p(\Phi_-) + \check{R}_p(\bar{\Phi}_-).$$

provided p, d satisfy the conditions of the proposition. \square

8.4 Notes and references

1. The memoryless channel model is the simplest, yet is basic in information theory. Already for this case, the coding theorems are not at all trivial. At the same time they give a basis for considering more complicated and realistic scenarios, taking into account memory effects. Moreover, the effects of entanglement are demonstrated in the memoryless case in a most spectacular way. Therefore, in our presentation throughout the book we concentrate on the memoryless configurations. Proposition 8.2, proved in the paper [98], was later generalized to memory channels by Kretschmann and Werner [141]. A generalization to arbitrary channels in the spirit of Han and Verdú's information spectrum approach was given by Hayashi and Nagaoka [79], [78].

2. The fruitful connection between the χ -capacity and the relative entropy, in particular the maximal distance property, was observed by Schumacher and Westmoreland [179].

3. One of the first mentions of the additivity problem was in the paper of Bennett, Fuchs and Smolin [22]. The formulation in terms of the four capacities was given by Bennett and Shor [23]. Remarkably, as shown by Shor [192], there is an intermediate capacity, which is obtained by allowing adaptive measurement on the output, and which lies strictly between $C_{1,1}$ and $C_{1,\infty}$ for the "lifted trines" channel.

Theorem 8.17 consists of several implications, some of which are individual, i.e. hold for any fixed pair of channels, while other only hold globally, e.g. "if the minimal output entropy is additive for all channels, so is the (constrained) χ -capacity". Matsumoto, Shimono, and Winter [157] observed the relation between a constrained version of the χ -capacity and the entanglement of formation. Audenaert and Braunstein [13] demonstrated the relevance of the methods of convex analysis, while the most profound global implications were established by Shor, who developed an ingenious extension construction that allows one to strengthen the individual channel's properties. See the paper of Shor [193] for a detailed account as well as for further references. Additional information on the equivalences of various additivity properties can be found in Chapter 9 of Hayashi's book [78]. The fact that global validity of the property (8.23) implies additivity (8.22) for all pairs of channels Φ_1, Φ_2 , (and a similar implication for the minimal output entropy) was proved by Fukuda and Wolf [65].

The relation between the additivity properties of complementary channels was noticed and studied by Holevo [103] and by King, Matsumoto, Natanson, and Ruskai [132]. Proposition 8.23 is proved in the paper of Holevo and Shirokov [109], where a mathematical description was given to the channel extension construction, and the super-additivity conjecture (8.36) and its equivalence to the additivity of constrained χ -capacity were studied systematically.

The Lindblad–Ozawa inequality was established in its final form in [162]. The validity of the additivity conjecture for entanglement-breaking channels was established

by Shor [190]. Other notable results include King's proofs of conjectures (8.22) and (8.33) for unital qubit [129] and depolarizing channels, as well as their complementary channels [103], [132]. Most of these results can be derived by proving the additivity of the minimal output Rényi entropies (8.53) for all values of p . A significant role in these proofs is played by the Lieb–Thirring inequality, see [27], Section X.2. For a detailed account of these and other partial results on the additivity problem, see the survey [104].

The counterexample of a transpose-depolarizing channel was proposed by Werner and Holevo [216]. A breakthrough came with the results of Winter [225] and Hayden [80], who had shown that the additivity of the minimal Rényi entropy with parameter values $p > 2$ and $1 < p \leq 2$, correspondingly, does not hold in sufficiently high dimensions. The phenomenon of entropy concentration was explored in [82], basing on a noncommutative generalization of Bernstein–Chernoff–Hoeffding inequality. Basing himself on the methods of these works, Hastings [76] considered mixtures of random unitary operators and sketched a proof of the violation of additivity for the minimal von Neumann entropy for large n , n/d . By Theorem 8.17, this implies the existence of counterexamples to all forms of the additivity, as well as to equality (8.24), as follows from [65]. The original argument of Hastings was improved and simplified in the subsequent works of Fukuda, King, and Moser [64], and Brandão and M. Horodecki [30]. Aubrun, Szarek, and E. Werner [11] pointed out a profound connection of these results to Dvoretzky–Milman's Theorem in the asymptotic theory of finite-dimensional Banach spaces. The constructive counterexample to the additivity of minimum output Rényi entropy of quantum channels for $p > 2$ is due to Grudka, M. Horodecky, and Pankowski [73].

An important corollary of these negative results is that, in general, the classical capacity $C(\Phi) \neq C_X(\Phi)$. By definition, $C(\Phi^{\otimes n}) = nC(\Phi)$. However, whether the classical capacity $C(\Phi)$ can be nonadditive for a pair of different quantum channels Φ_1, Φ_2 remains an open question.

Chapter 9

Entanglement-assisted classical communication

9.1 The gain of entanglement assistance

The construction of the superdense coding protocol (Section 3.3.3) allows for generalization to the case of non-ideal channel Φ between arbitrary finite dimensional quantum systems A and A' . Consider the following protocol of the classical information transmission through the channel Φ . Systems A (transmitter) and B (receiver) share an entangled (pure) state S_{AB} , which is distributed to them with a procedure of the kind described in Section 3.2.1. We shall assume that $\dim \mathcal{H}_A \leq \dim \mathcal{H}_B$. System A does some encoding of classical messages i , arriving with probabilities π_i into operations (channels) \mathcal{E}_A^i , acting in \mathcal{H}_A . These operations are applied by A to its part of the state S_{AB} , so that the state of the system AB is transformed into $(\mathcal{E}_A^i \otimes \text{Id}_B)[S_{AB}]$. After that, A sends its part of this state through the channel Φ , thus driving the whole system AB into the state $(\Phi \circ \mathcal{E}_A^i \otimes \text{Id}_B)[S_{AB}]$, accessible to B . Next, B makes a measurement in the space $\mathcal{H}_{A'B}$ in order to extract the classical information. In fact, block encoding is allowed, so that the whole of this description refers to the n -th tensor degree of $\mathcal{H}_{A'B}$. We are interested in the classical capacity of this protocol, which is called the *entanglement-assisted classical capacity* of the channel Φ .

The maximum over measurements of B can be evaluated using the Coding Theorem 5.4 for the classical capacity. First, we introduce the quantity

$$C_{ea}^{(1)}(\Phi) = \sup_{\pi_i, \mathcal{E}_A^i, S_{AB}} \chi \left(\{\pi_i\}; (\Phi \otimes \text{Id}_B)[S_{AB}^i] \right), \quad (9.1)$$

which takes into account measurements in the system $A'B$. Using the channel n times and allowing arbitrary collective (entangled) measurement on B 's side, one gets

$$C_{ea}^{(n)}(\Phi) = C_{ea}^{(1)}(\Phi^{\otimes n}). \quad (9.2)$$

Now, the full entanglement-assisted classical capacity is

$$C_{ea}(\Phi) = \lim_{n \rightarrow \infty} \frac{1}{n} C_{ea}^{(1)}(\Phi^{\otimes n}). \quad (9.3)$$

The following theorem gives a simple “one-letter” expression for $C_{ea}(\Phi)$ in terms of quantum mutual information,

Theorem 9.1 (Bennett, Shor, Smolin and Thapliyal [24]). *The entanglement-assisted classical capacity of the channel Φ is equal to*

$$C_{ea}(\Phi) = \max_{S_A} I(S_A, \Phi) = \max_{S_A} [H(S_A) + H(\Phi(S_A)) - H(S_A; \Phi)]. \quad (9.4)$$

The proof of this theorem will be given in the following sections. Here, we wish to compare entanglement-assisted and (unassisted) classical capacities.

Example 9.2. Consider the quantum erasure channel Φ_p . We have

$$H(\Phi_p[S_A]) = (1 - p)H(S_A) + h_2(p).$$

Using the fact that the complementary channel $\tilde{\Phi}_p = \Phi_{1-p}$ (see Exercise 6.35) together with Definition 7.28 of the entropy exchange, we have $H(S_A; \Phi) = H(\tilde{\Phi}[S_A]) = pH(S_A) + h_2(p)$. Hence,

$$I(S_A, \Phi_p) = H(S_A) + (1 - p)H(S_A) - pH(S_A) = 2(1 - p)H(S_A),$$

whence $C_{ea}(\Phi_p) = 2(1 - p) \log d$, which is as twice as large as the classical capacity of the erasure channel.

Proposition 9.3. *If the channel Φ is covariant, the maximum in (9.4) is attained for an invariant state S_A . In particular, if Φ is irreducibly covariant, it is attained on the chaotic state.*

Proof. Let the channel Φ satisfy the covariance condition (6.48). In this case,

$$I(V_g^A S_A V_g^{A*}, \Phi) = I(S_A, \Phi). \quad (9.5)$$

This follows from the expression (7.52) for the mutual information, unitary covariance of the input and output entropies, and the corresponding property of the entropy exchange,

$$H(V_g^A S_A V_g^{A*}, \Phi) = H(S_A, \Phi). \quad (9.6)$$

To prove this last property, consider formula (7.47) for the entropy exchange and note that purification of the state $V_g^A S_A V_g^{A*}$ is given by the vector $(V_g^A \otimes I_R)|\psi_{AR}\rangle$. Substituting this into (7.47) and again using the covariance of Φ and the unitary invariance of the entropy gives (9.6).

Now, let S_A be an optimal state for the channel Φ . Consider the V_g^A -invariant state

$$\bar{S} = \frac{1}{|G|} \sum_{g \in G} V_g^A S_A V_g^{A*}.$$

By Proposition 7.31, the function $S \rightarrow I(S, \Phi)$ is concave, and therefore

$$I(\bar{S}, \Phi) \geq \frac{1}{|G|} \sum_{g \in G} I(V_g^A S_A V_g^{A*}, \Phi) = I(S_A, \Phi),$$

where the last equality follows from (9.5). \square

Example 9.4. Consider the depolarizing channel (6.49). By using unitary covariance, one sees that $I(\bar{S}, \Phi)$ achieves its maximum at the chaotic state $\bar{S} = \frac{I}{d}$, for which $H(\bar{S}) = H(\Phi[\bar{S}]) = \log d$. The entropy exchange $H(\bar{S}, \Phi)$ is given by the relation (7.51) whence

$$C_{ea}(\Phi) = \log d^2 + \left(1 - p \frac{d^2 - 1}{d^2}\right) \log \left(1 - p \frac{d^2 - 1}{d^2}\right) + p \frac{d^2 - 1}{d^2} \log \frac{p}{d^2}. \quad (9.7)$$

This should be compared with the unassisted classical capacity C , which is equal to C_χ , given by the formula (8.19).

One can see, in particular, that the *gain of entanglement assistance* $C_{ea}/C \rightarrow d+1$ in the limit of strong noise $p \rightarrow 1$ (note that both capacities tend to zero!) Moreover, taking the maximal possible value $p = \frac{d^2}{d^2-1}$, we obtain

$$C_{ea} = \log \frac{d^2}{d^2-1},$$

$$C = C_\chi = \frac{1}{d+1} \log \frac{d}{d+1} + \frac{d}{d+1} \log \frac{d^2}{d^2-1}.$$

The ratio C_{ea}/C increases monotonically from the value 5.08 for $d = 2$, tightly approaching the asymptotic line $2(d+1)$ as d grows. Recall (Proposition 6.40) that the depolarizing channel is entanglement-breaking for all $p \geq d/(d+1)$. Nevertheless, $C_{ea} > C$ also in this case. In the next section we will show that this inequality holds generically even for q-c channels.

Here, entanglement again appears as a “catalyst” of the transmission of classical information through the quantum channel – it may indefinitely increase the classical capacity of noisy channels, while being unable to transmit the information on its own.

Exercise 9.5. Show that for a qubit unital channel of the form (6.59) with Λ given by (6.64),

$$C_{ea}(\Phi) = I(\bar{S}, \Phi) = 2 \log 2 + \sum_{\gamma=0,x,y,z} \mu_\gamma \log \mu_\gamma.$$

Hint: The entropy exchange $H(\bar{S}, \Phi)$ can be computed by using Proposition 7.29 and the Kraus representation (6.64), resulting in

$$H(\bar{S}, \Phi) = - \sum_{\gamma=0,x,y,z} \mu_\gamma \log \mu_\gamma.$$

Compare this with the capacity $C_\chi(\Phi)$, given by formula (8.21).

In general, there are simple inequalities relating entanglement-assisted and unassisted classical capacities:

$$C_\chi(\Phi) \leq C_{ea}(\Phi) \leq \log d + C_\chi(\Phi). \quad (9.8)$$

Take S_A to be a mixture of arbitrary pure states S_j with probabilities p_j . In this case, (7.50) implies

$$I(S_A, \Phi) \leq H\left(\sum_j p_j S_j\right) + H\left(\sum_j p_j \Phi(S_j)\right) - \sum_j p_j H\left(\Phi(S_j)\right).$$

Taking the maxima in both sides and using (9.4), we obtain the second inequality in (9.8).

Although the entanglement-assisted protocol uses additional resources compared to unassisted transmission of classical information, the first inequality in (9.8) is not quite obvious, because of the special nature of the encoding procedure used by A . To avoid this, let us show that the definition of $C_{ea}^{(1)}(\Phi)$, and hence of $C_{ea}(\Phi)$, can be formulated without explicit introduction of the encoding operations \mathcal{E}_A^i , namely,

$$C_{ea}^{(1)}(\Phi) = \sup_{\pi_i, \{S_{AB}^i\} \in \Sigma_B} \left[H\left(\sum_i \pi_i (\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) - \sum_i \pi_i H\left((\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) \right], \quad (9.9)$$

where Σ_B is the collection of families of the states $\{S_{AB}^i\}$ that satisfy the condition that their partial states S_B^i may not depend on i , $S_B^i = S_B$.

The first inequality in (9.8) then follows by taking $S_{AB}^i = S_A^i \otimes S_B$. To prove (9.9), it is sufficient to establish the following fact.

Lemma 9.6. *Let $\{S_{AB}^i\}$ be a family of the states satisfying the condition $S_B^i = S_B$. Then there exist a pure state S_{AB} and encodings \mathcal{E}_A^i such that*

$$S_{AB}^i = (\mathcal{E}_A^i \otimes \text{Id}_B)[S_{AB}]. \quad (9.10)$$

Proof. For simplicity assume first that S_B is nondegenerate. In this case,

$$S_B = \sum_{k=1}^d \lambda_k |e_k^B\rangle \langle e_k^B|,$$

where $\lambda_k > 0$ and $\{|e_k^B\rangle\}$ is an orthonormal basis in \mathcal{H}_B . Let $\{|e_k^A\rangle\}$ be an orthonormal basis in \mathcal{H}_A . For a vector $|\psi^A\rangle = \sum_{k=1}^d c_k |e_k^A\rangle$ we denote $|\bar{\psi}^B\rangle = \sum_{k=1}^d \bar{c}_k |e_k^B\rangle$. The map $|\psi^A\rangle \rightarrow |\bar{\psi}^B\rangle$ is an anti-isomorphism of \mathcal{H}_A and \mathcal{H}_B . Set

$$|\psi_{AB}\rangle = \sum_{k=1}^d \sqrt{\lambda_k} |e_k^A\rangle \otimes |e_k^B\rangle,$$

so that $S_{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$, and define encodings by the relation

$$\mathcal{E}_A^i \left[|\psi^A\rangle\langle\phi^A| \right] = \langle\bar{\psi}^B| S_B^{-1/2} S_{AB}^i S_B^{-1/2} |\bar{\phi}^B\rangle, \quad |\psi^A\rangle, |\phi^A\rangle \in \mathcal{H}_A.$$

In this case, one can check that \mathcal{E}_A^i are indeed channels fulfilling the formula (9.10).

In the case where S_B is degenerate, the above construction should be modified by replacing $S_B^{-1/2} S_{AB}^i S_B^{-1/2}$ in the above formula with $\sqrt{S_B^-} S_{AB}^i \sqrt{S_B^-} + P_B^0$, where S_B^- is the generalized inverse of S_B and P_B^0 is the projection onto the null subspace of S_B . \square

9.2 The classical capacities of quantum observables

In this section, we wish to compare the classical capacities C , C_{ea} for the special class of q-c channels, when it is possible to establish a necessary and sufficient condition for their coincidence. Let $M = \{M_y; y \in \mathcal{Y}\}$ be a quantum observable in a Hilbert space \mathcal{H} . Consider the measurement channel

$$\mathcal{M}[S] = \sum_y |e_y\rangle\langle e_y| \text{Tr} S M_y. \quad (9.11)$$

Since any q-c channel is entanglement-breaking, its classical capacity is given by the one-letter expression

$$C(\mathcal{M}) = C_\chi(\mathcal{M}) = \max_\pi \mathcal{I}(\pi; M),$$

where $\pi = \{\pi_x, S_x\}$ is an ensemble that assigns the probabilities π_x to states S_x and

$$\mathcal{I}(\pi; M) = H(P_{\bar{S}_\pi}) - \sum_x \pi_x H(P_{S_x})$$

is the Shannon information between the input x and the output y . Here, $\bar{S}_\pi = \sum_x \pi_x S_x$, $P_S = \{\text{Tr} S M_y\}$ is the probability distribution of the measurement outcomes and $H(P)$ is the Shannon entropy of a probability distribution P . This can be rewritten as

$$C(\mathcal{M}) = \max_S \left[H(P_S) - \min_{\pi: \bar{S}_\pi = S} \sum_x \pi_x H(P_{S_x}) \right], \quad (9.12)$$

where the minimum is over input ensembles π with the fixed average state $\bar{S}_\pi = S$.

Next, consider the entanglement-assisted capacity which, according to Theorem 9.1, is given by the formula

$$C_{ea}(\mathcal{M}) = \max_S I(S; \mathcal{M}), \quad (9.13)$$

where

$$I(S; \mathcal{M}) = H(S) + H(\mathcal{M}[S]) - H(S, \mathcal{M})$$

is the quantum mutual information. Here, $H(\cdot)$ is the von Neumann entropy and $H(S, \mathcal{M})$ is the entropy exchange. Let $p_y = \text{Tr } S M_y$ and V_y be an operator satisfying $M_y = V_y^* V_y$, for example, $V_y = M_y^{1/2}$. Then the density operator $\frac{V_y S V_y^*}{p_y} = S(y|M)$ can be interpreted as posterior state of the measurement of observable M with instrument $S \rightarrow \{V_y S V_y^*\}$ in state S . We use the following formula [186]:

$$I(S; \mathcal{M}) = H(S) - \sum_y (\text{Tr } S M_y) H(S(y|M)). \quad (9.14)$$

Indeed, writing the Kraus decomposition

$$\mathcal{M}[S] = \sum_{y,j} |e_y\rangle \langle h_j| V_y S V_y^* |h_j\rangle \langle e_y|$$

and applying Proposition 7.29, we obtain

$$\begin{aligned} H(S, \mathcal{M}) &= H\left(\sum_y V_y S V_y^* \otimes |e_y\rangle \langle e_y|\right) \\ &= H\left(\sum_y p_y S(y|M) \otimes |e_y\rangle \langle e_y|\right) \\ &= H(P_S) + \sum_y p_y H(S(y|M)) \end{aligned}$$

while $H(\mathcal{M}[S]) = H(P_S)$, hence (9.14). The entanglement-assisted classical capacity of the channel \mathcal{M} follows by substituting this expression into (9.13).

Let us now describe the *ensemble-observable duality*. If $M = \{M_y; y \in \mathcal{Y}\}$ is a quantum observable and $\pi = \{p_x, S_x; x \in \mathcal{X}\}$ an ensemble of quantum states, then

$$p_{xy} = p_x \text{Tr } S_x M_y$$

is a probability distribution on $\mathcal{X} \times \mathcal{Y}$. On the other hand,

$$p_{xy} = p'_y \text{Tr } S'_y M'_x,$$

where, denoting $\bar{S}_\pi = \sum_x p_x S_x$, we have $p'_y S'_y = \bar{S}_\pi^{1/2} M_y \bar{S}_\pi^{1/2}$ so that $p'_y = \text{Tr } \bar{S}_\pi M_y$ and $M'_x = p_x \bar{S}_\pi^{-1/2} S_x \bar{S}_\pi^{-1/2}$. Here $M'_\pi = \{M'_x; x \in \mathcal{X}\}$ is the new observable, and $\pi' = \{p'_y, S'_y; y \in \mathcal{Y}\}$ is the new ensemble. Therefore, the Shannon information between x, y is

$$\mathcal{I}(\pi, M) = \mathcal{I}(\pi', M'_\pi).$$

Introducing the ensemble $\pi'_S = \left\{ \text{Tr } S M_y, \frac{S^{1/2} M_y S^{1/2}}{\text{Tr } S M_y} \right\}$, for a given state S , we have the following duality relations,

Proposition 9.7. Denoting $A(\pi'_S) = \max_{M''} \mathcal{I}(\pi'_S, M'')$ the accessible information and $\chi(\pi'_S) = H(\sum_y p'_y S'_y) - \sum_y p'_y H(S'_y)$ the χ -quantity of the dual ensemble π'_S , we have

$$C(\mathcal{M}) \equiv \max_{\pi} \mathcal{I}(\pi, M) = \max_S A(\pi'_S), \quad (9.15)$$

$$C_{ea}(\mathcal{M}) \equiv \max_S I(S; \mathcal{M}) = \max_S \chi(\pi'_S). \quad (9.16)$$

Proof. Relation (9.15) follows from

$$\begin{aligned} \max_{\pi} \mathcal{I}(\pi, M) &= \max_S \max_{\pi: \bar{S}_{\pi}=S} \mathcal{I}(\pi, M) \\ &= \max_S \max_{M'_{\pi}: \bar{S}_{\pi}=S} \mathcal{I}(\pi'_S, M'_{\pi}) \\ &= \max_S \max_{M''} \mathcal{I}(\pi'_S, M'') \\ &= \max_S A(\pi'_S), \end{aligned}$$

where in the third equality we used the fact that for any observable $M'' = \{M_y\}$ there is an ensemble $\pi = \left\{ \frac{\text{Tr } S M_y}{d}, \frac{M_y}{\text{Tr } S M_y} \right\}$ such that $M'' = M'_{\pi}$ for this ensemble.

Notice that $\sum_y p'_y S'_y = S$, and $H(S'_y) = H\left(\frac{V_y S V_y^*}{p'_y}\right)$, where V_y is an arbitrary operator that satisfies $M_y = V_y^* V_y$, because the operators $V_y S V_y^* = V_y S^{1/2} S^{1/2} V_y^*$ and $S^{1/2} V_y^* V_y S^{1/2} = S^{1/2} M_y S^{1/2} = p'_y S'_y$ are unitarily equivalent via polar decomposition and hence have the same spectrum. The density operator $\frac{V_y S V_y^*}{p'_y} = S(y|M)$ is the posterior state of the measurement of the observable M with the instrument $\{V_y\}$ in the state S . Thus, the χ -quantity of the dual ensemble is

$$\chi(\pi'_S) = H(S) - \sum_y p'_y H(S(y|M)) = I(S; \mathcal{M}) \quad (9.17)$$

and hence, in addition to (9.15), we have, via (9.14), the duality relation (9.16). \square

Let us now recall the bound of Theorem 5.9

$$A(\pi) \leq H\left(\sum_y p_y S_y\right) - \sum_y p_y H(S_y) \equiv \chi(\pi) \quad (9.18)$$

with the equality attained if and only if the operators $p_y S_y$ all commute. Applying this to the dual situation, we obtain

$$A(\pi'_S) \leq H\left(\sum_y p'_y S'_y\right) - \sum_y p'_y H(S'_y) = \chi(\pi'_S).$$

In this case, the necessary and sufficient condition for the equality becomes

$$S^{1/2} M_y S M_{y'} S^{1/2} = S^{1/2} M_{y'} S M_y S^{1/2} \quad (9.19)$$

for all y, y' . Therefore, by (9.15), (9.16), the necessary and sufficient condition for the equality $C_{ea}(\mathcal{M}) = C(\mathcal{M})$ is that condition (9.19) is fulfilled for a density operator S maximizing the quantity (9.17). In particular, in the case of a covariant observable $M_g = V_g M_0 V_g^*$, where V_g is an irreducible representation of a symmetry group G , Proposition 9.3 implies $S = I/d$ and condition (9.19) reduces to the fact that all components of the observable commute, i.e. it is essentially a classically randomized observable.

Consider the case of an overcomplete system $M_y = |\psi_y\rangle\langle\psi_y|$ in d -dimensional Hilbert space \mathcal{H} , where $S = I/d$. The corresponding ensemble of pure states is $\pi'_S \equiv \bar{\pi} = \left\{ \frac{\langle\psi_y|\psi_y\rangle}{d}, \frac{|\psi_y\rangle\langle\psi_y|}{\langle\psi_y|\psi_y\rangle} \right\}$ and $\chi(\bar{\pi}) = \log d = C_{ea}(\mathcal{M})$ (notice that this is also equal to the classical capacity of the c-q channel $y \rightarrow \frac{|\psi_y\rangle\langle\psi_y|}{\langle\psi_y|\psi_y\rangle}$, since this is the maximal possible value). Condition (9.19) amounts to

$$|\psi_y\rangle\langle\psi_y|\psi_{y'}\rangle\langle\psi_{y'}| = |\psi_{y'}\rangle\langle\psi_{y'}|\psi_y\rangle\langle\psi_y|.$$

We can always assume that the vectors $|\psi_y\rangle$ are all pairwise linearly independent. In this case, the last condition is equivalent to the fact that they form an orthonormal basis.

Thus we conclude that $C_{ea}(\mathcal{M}) > C(\mathcal{M})$, unless $\{|\psi_y\rangle\}$ is an orthonormal basis.

Now, let Θ be the unit sphere in \mathcal{H} and let $\nu(d\theta)$ be the uniform distribution on Θ . To avoid notational confusion with the differential, we denote in the argument below the dimensionality with $d_A = \dim \mathcal{H}$. According to relation (6.55), the family of unit vectors $\{|\theta\rangle; \theta \in \Theta\}$ can be considered a continuous overcomplete system. The corresponding measurement channel \mathcal{M} maps the density operator S to the probability distribution on Θ :

$$\mathcal{M} : S \rightarrow d_A \langle \theta | S | \theta \rangle \nu(d\theta). \quad (9.20)$$

It can be shown that, similar to the discrete overcomplete systems considered above, $C_{ea}(\mathcal{M}) = \log d_A$. Let us show that

$$C(\mathcal{M}) = \log d_A - \log e \sum_{k=2}^{d_A} \frac{1}{k}. \quad (9.21)$$

Thus, for $d_A \rightarrow \infty$ we have $C(\mathcal{M}) \rightarrow \log e (1 - \gamma)$, where $\gamma \approx 0.577$ is Euler's constant. At the same time, $C_{ea}(\mathcal{M}) = \log d_A \rightarrow \infty$.

All outcome probability distributions (9.20) are absolutely continuous with respect to $\nu(d\theta)$, and hence we can use the differential entropy

$$h(p_S) = - \int_{\Theta} p_S(\theta) \log p_S(\theta) \nu(d\theta),$$

where $p_S(\theta) = d_A \langle \theta | S | \theta \rangle$, to get the continuous analog of formula (9.12)

$$C(\mathcal{M}) = C_X(\mathcal{M}) = \max_S \left[h(p_S) - \min_{\pi: S_\pi = S} \sum_x \pi_x h(p_{S_x}) \right]. \quad (9.22)$$

The channel \mathcal{M} is covariant with respect to the irreducible action of the unitary group, in the sense that

$$\mathcal{M}(USU^*) = d_A \langle U^* \theta | S | U^* \theta \rangle \nu(d\theta).$$

Therefore, similar to (8.15)

$$C_X(\mathcal{M}) = h\left(\mathcal{M}\left(\frac{I}{d_A}\right)\right) - \min_{\theta'} h(\mathcal{M}(|\theta'\rangle\langle\theta'|)). \quad (9.23)$$

But $\mathcal{M}\left(\frac{I}{d_A}\right)$ is the uniform distribution over Θ , with density $p(\theta) \equiv 1$. Hence, $h\left(\mathcal{M}\left(\frac{I}{d_A}\right)\right) = 0$. On the other hand,

$$-h(\mathcal{M}(|\theta'\rangle\langle\theta'|)) = \int_{\Theta} d_A |\langle \theta | \theta' \rangle|^2 \log \left[d_A |\langle \theta | \theta' \rangle|^2 \right] \nu(d\theta). \quad (9.24)$$

By the unitary invariance of ν , this quantity is the same for all θ' . Hence, there is no need for minimization in (9.23). For its computation, we use relation (6.57). With this, (9.24) becomes

$$-\int_0^1 (d_A r^2) \log(d_A r^2) d(1-r^2)^{d_A-1} = \int_0^1 [d_A(1-u)] \log[d_A(1-u)] d(u^{d_A-1}),$$

where $u = 1 - r^2$, which, after integrations by parts, gives (9.21).

9.3 Proof of the Converse Coding Theorem

Now, we finally come to the proof of Theorem 9.1.

To prove the inequality

$$C_{ea}(\Phi) \leq \max_{S_A} I(S_A, \Phi), \quad (9.25)$$

we first prove that

$$C_{ea}^{(1)}(\Phi) \leq \max_{S_A} I(S_A, \Phi). \quad (9.26)$$

Let us denote by \mathcal{E}_A^i the encodings used by A . Let S_{AB} be the pure state, initially shared between A and B . In this case, the state of the system AB (resp. A) after the encoding is

$$S_{AB}^i = (\mathcal{E}_A^i \otimes \text{Id}_B)[S_{AB}], \quad S_A^i = \mathcal{E}_A^i[S_A]. \quad (9.27)$$

Note that the partial state of B does not change after the encoding, $S_B^i = S_B$. We will prove that

$$H\left(\sum_i \pi_i (\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) - \sum_i \pi_i H\left((\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) \leq I\left(\sum_i \pi_i S_A^i; \Phi\right). \quad (9.28)$$

The maximum of the left hand side with respect to all possible π_i , \mathcal{E}_A^i is just $C_{ea}^{(1)}(\Phi)$, from which (9.26) follows.

By using the subadditivity of the quantum entropy, we can evaluate the first term in the left hand side of (9.28) as

$$H\left(\sum_i \pi_i \Phi[S_A^i]\right) + H(S_B) = H\left(\Phi\left[\sum_i \pi_i S_A^i\right]\right) + \sum_i \pi_i H(S_B).$$

Here, the first term already provides us with the output entropy from $I(\sum_i \pi_i S_A^i; \Phi)$. Let us proceed with the evaluation of the remainder

$$\sum_i \pi_i \left[H(S_B) - H\left((\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) \right].$$

We first show that the term in squared brackets does not exceed $H(S_A^i) - H((\Phi \otimes \text{Id}_{R^i})[S_{AR^i}^i])$, where R^i is the purifying (reference) system for S_A^i , and $S_{AR^i}^i$ is the purified state. To this end consider the unitary extension of the encoding \mathcal{E}_A^i with the environment E_i , which is initially in a pure state, according to Theorem 6.18. From (9.27) we see that we can take $R^i = BE_i$ (after the unitary interaction, which involves only AE_i). In this case, again denoting by primes states after the application of the channel Φ , we have

$$H(S_B) - H\left((\Phi \otimes \text{Id}_B)[S_{AB}^i]\right) = H(S_B) - H(S_{A'B}^i) = -H_i(A'|B), \quad (9.29)$$

where the lower index i of the conditional entropy refers to the joint state $S_{A'B}^i$. Similarly,

$$\begin{aligned} H(S_A^i) - H\left((\Phi \otimes \text{Id}_{R^i})[S_{AR^i}^i]\right) &= H(S_{R^i}^i) - H(S_{A'R^i}^i) \\ &= -H_i(A'|R^i) = -H_i(A'|BE_i), \end{aligned}$$

which is greater than or equal to (9.29), by the monotonicity of the conditional entropy.

Using the concavity of the function $S_A \rightarrow H(S_A) - H((\Phi \otimes \text{Id}_R)[S_{AR}])$, to be shown below, we get

$$\sum_i \pi_i \left[H(S_A^i) - H((\Phi \otimes \text{Id}_R)[S_{AR}^i]) \right] \leq H\left(\sum_i \pi_i S_A^i\right) - H\left((\Phi \otimes \text{Id}_R)[\hat{S}_{AR}]\right),$$

where \hat{S}_{AR} is the state purifying $\sum_i \pi_i S_A^i$ with a reference system R .

To complete this proof, it remains to show the above concavity. By introducing the environment E for the channel Φ , we have

$$\begin{aligned} H(S_A) - H((\Phi \otimes \text{Id}_R)[S_{AR}]) &= H(S_R) - H(S_{A'R}) \\ &= H(S_{A'E'}) - H(S_{E'}) = H(A'|E') \end{aligned}$$

Now, $H(A'|E')$ is a concave function of $S_{A'E'}$ by Corollary 7.27. The map $S_A \rightarrow S_{A'E'}$ is affine and therefore $H(A'|E')$ is a concave function of S_A .

Applying the same argument to the channel $\Phi^{\otimes n}$ gives

$$C_{ea}^{(n)}(\Phi) \leq \max_{S_A} I(S_A, \Phi^{\otimes n}). \quad (9.30)$$

Then, from the subadditivity of the quantum mutual information (property iii. in Proposition 7.31) we have

$$\max_{S_{12}} I(S_{12}, \Phi_1 \otimes \Phi_2) = \max_{S_1} I(S_1, \Phi_1) + \max_{S_2} I(S_2, \Phi_2),$$

whence the remarkable additivity follows

$$\max_{S_{A^n}} I(S_{A^n}, \Phi^{\otimes n}) = n \max_{S_A} I(S_A, \Phi).$$

Therefore, we finally obtain (9.25). \square

9.4 Proof of the Direct Coding Theorem

Here, we prove the inequality

$$C_{ea}(\Phi) \geq \max_{S_A} I(S_A, \Phi). \quad (9.31)$$

First, let us show by generalizing the dense coding protocol, that

$$C_{ea}^{(1)}(\Phi^{\otimes n}) \geq I\left(\frac{P}{\dim P}, \Phi^{\otimes n}\right) \quad (9.32)$$

for an arbitrary projection P in $\mathcal{H}_A^{\otimes n}$.

Indeed, let $P = \sum_{k=1}^m |e_k\rangle\langle e_k|$, where $\{e_k; k = 1, \dots, m = \dim P\}$ is an orthonormal system. Consider the discrete Weyl system $\{W_{\alpha\beta}; \alpha, \beta = 0, \dots, m-1\}$, defined by the relations (6.50) on the subspace $\mathcal{H}_P = P\mathcal{H}_A^{\otimes n}$.

Since $\dim \mathcal{H}_A \leq \dim \mathcal{H}_B$, we can assume that $\mathcal{H}_A \subseteq \mathcal{H}_B$. Let

$$|\psi_{AB}\rangle = \frac{1}{\sqrt{m}} \sum_{k=1}^m |e_k\rangle \otimes |e_k\rangle.$$

Exercise 9.8. Show that the system $\{(W_{\alpha\beta} \otimes I_B) |\psi_{AB}\rangle; \alpha, \beta = 0, \dots, m-1\}$ is an orthonormal basis in $\mathcal{H}_P \otimes \mathcal{H}_P$. In particular,

$$\sum_{\alpha, \beta=0}^{m-1} (W_{\alpha\beta} \otimes I_B) |\psi_{AB}\rangle \langle \psi_{AB}| (W_{\alpha\beta} \otimes I_B)^* = P \otimes P. \quad (9.33)$$

The operators $W_{\alpha\beta}$ will play a role similar to the Pauli matrices in the dense coding protocol for qubits. Take the classical signal to be transmitted as $i = (\alpha, \beta)$, with equal probabilities $1/m^2$, the entangled state $|\psi_{AB}\rangle\langle\psi_{AB}|$, and the unitary encodings $\mathcal{E}_A^i[S] = W_{\alpha\beta} S W_{\alpha\beta}^*$. Now, we have

$$C_{ea}^{(1)}(\Phi^{\otimes n}) \geq H\left(\frac{1}{m^2} \sum_{\alpha, \beta} (\Phi \otimes \text{Id}_B)^{\otimes n} [S_{AB}^{\alpha\beta}]\right) - \frac{1}{m^2} \sum_{\alpha, \beta} H\left((\Phi \otimes \text{Id}_B)^{\otimes n} [S_{AB}^{\alpha\beta}]\right),$$

where $S_{AB}^{\alpha\beta} = (W_{\alpha\beta} \otimes I_B^{\otimes n}) |\psi_{AB}\rangle\langle\psi_{AB}| (W_{\alpha\beta} \otimes I_B^{\otimes n})^*$. By (9.33) the first term in the right hand side is equal to $H((\Phi \otimes \text{Id}_B)[\frac{P}{m} \otimes \frac{P}{m}]) = H(\frac{P}{m}) + H(\Phi[\frac{P}{m}])$. Since $S_{AB}^{\alpha\beta}$ is a purification of $\frac{P}{m}$, the entropies in the second term are all equal to $H(\frac{P}{m}, \Phi)$. By the expression for the quantum mutual information

$$I(S_A, \Phi) = H(S_A) + H(\Phi[S_A]) - H(S_A; \Phi)$$

with $S_A = \frac{P}{m}$, this proves (9.32). For future use, note that the last term in the quantum mutual information – the entropy exchange – is equal to the final environment entropy $H(S'_E) = H(\Phi[S_A])$, where Φ is the complementary channel from the system space \mathcal{H}_A to the environment space \mathcal{H}_E .

Now let $S_A = S$ be an arbitrary state in \mathcal{H}_A and let $\hat{P}^{n,\delta}$ be the *strongly typical projector* of the state $S^{\otimes n}$ in the space $\mathcal{H}_A^{\otimes n}$, defined below. We shall prove that for an *arbitrary* channel Ψ from \mathcal{H}_A to a possibly other Hilbert space $\tilde{\mathcal{H}}$ the following relation holds

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} H \left(\Psi^{\otimes n} \left[\frac{\hat{P}^{n,\delta}}{\dim \hat{P}^{n,\delta}} \right] \right) = H(\Psi[S]).$$

This would imply, by the above expressions for the mutual information and the entropy exchange, that

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} I \left(\frac{\hat{P}^{n,\delta}}{\dim \hat{P}^{n,\delta}}, \Phi^{\otimes n} \right) = I(S, \Phi), \quad (9.34)$$

and hence, by (9.32), the required inequality (9.31).

Definition 9.9. Let us fix small positive δ , and let λ_j be the eigenvalues, $|e_j\rangle$ the eigenvectors of the density operator S . In this case, the eigenvalues and eigenvectors of $S^{\otimes n}$ are $\lambda_J = \lambda_{j_1} \cdots \lambda_{j_n}$, $|e_J\rangle = |e_{j_1}\rangle \otimes \cdots \otimes |e_{j_n}\rangle$, where $J = (j_1, \dots, j_n)$. The sequence J is called *strongly typical* if the numbers $N(j|J)$ of times that the symbol j appears in J satisfy the condition

$$\left| \frac{N(j|J)}{n} - \lambda_j \right| < \delta, \quad j = 1, \dots, d,$$

and $N(j|J) = 0$ if $\lambda_j = 0$. Let us denote the collection of all strongly typical sequences as $\hat{T}^{n,\delta}$. The *strongly typical projector* is defined as the following spectral projector of the density operator $S^{\otimes n}$:

$$\hat{P}^{n,\delta} = \sum_{J \in \hat{T}^{n,\delta}} |e_J\rangle \langle e_J|.$$

Let P^n be the probability distribution given by the eigenvalues λ_J . Then, by the Law of Large Numbers, $P^n(\hat{T}^{n,\delta}) \rightarrow 1$ as $n \rightarrow \infty$. For an arbitrary function $f(j)$, $j = 1, \dots, d$ and a sequence $J = (j_1, \dots, j_n) \in \hat{T}^{n,\delta}$

$$\left| \frac{f(j_1) + \cdots + f(j_n)}{n} - \sum_{j=1}^d \lambda_j f(j) \right| < \delta d \max f. \quad (9.35)$$

In particular, any strongly typical sequence is typical: taking $f(j) = -\log \lambda_j$ provides

$$n[H(S) - \delta_1] < -\log \lambda_J < n[H(S) + \delta_1], \quad (9.36)$$

where $\delta_1 = \delta d \max_{\lambda_j > 0} (-\log \lambda_j)$ (the converse is not true – not every typical sequence is strongly typical).

We shall need the following combinatorial result, the proof of which can be found in [44].

Exercise 9.10. Show that the size of the set $\hat{T}^{n,\delta}$ is estimated as

$$2^{n[H(S) - \Delta_n(\delta)]} < |\hat{T}^{n,\delta}| < 2^{n[H(S) + \Delta_n(\delta)]}, \quad (9.37)$$

where $H(S) = -\sum_{j=1}^d \lambda_j \log \lambda_j$, and $\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \Delta_n(\delta) = 0$.

Proof of the relation (9.34). Denote $d_{n,\delta} = \dim \hat{P}^{n,\delta} = |\hat{T}^{n,\delta}|$ and $\bar{S}^{n,\delta} = \frac{\hat{P}^{n,\delta}}{d_{n,\delta}}$. We will prove that

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} H(\Psi^{\otimes n} [\bar{S}^{n,\delta}]) = H(\Psi[S]) \quad (9.38)$$

for an arbitrary channel Ψ . We have

$$\begin{aligned} nH(\Psi[S]) - H(\Psi^{\otimes n} [\bar{S}^{n,\delta}]) &= H(\Psi[S]^{\otimes n}) - H(\Psi^{\otimes n} [\bar{S}^{n,\delta}]) \\ &= H(\Psi^{\otimes n} [\bar{S}^{n,\delta}]; \Psi^{\otimes n} [S^{\otimes n}]) \\ &\quad + \text{Tr} \log \Psi[S]^{\otimes n} (\Psi^{\otimes n} [\bar{S}^{n,\delta}] - \Psi[S]^{\otimes n}). \end{aligned} \quad (9.39)$$

Strictly speaking, this formula is correct if the density operator $\Psi[S]^{\otimes n}$ is nondegenerate, which we will assume for the moment. Later we shall show how the argument can be modified to the general case.

For the first term, by the monotonicity of the relative entropy, we have the estimate

$$H(\Psi^{\otimes n} [\bar{S}^{n,\delta}]; \Psi^{\otimes n} [S^{\otimes n}]) \leq H(\bar{S}^{n,\delta}; S^{\otimes n}),$$

with the right-hand side computed explicitly as

$$H(\bar{S}^{n,\delta}; S^{\otimes n}) = \sum_{J \in \hat{T}^{n,\delta}} \frac{1}{d_{n,\delta}} \log \frac{1}{d_{n,\delta} \lambda_J} = -\log d_{n,\delta} - \sum_{J \in \hat{T}^{n,\delta}} \frac{1}{d_{n,\delta}} \log \lambda_J,$$

which by (9.36), (9.37) is less than or equal to the quantity $n(\delta_1 + \Delta_n(\delta))$, converging to zero as $n \rightarrow \infty, \delta \rightarrow 0$.

By using the identity

$$\log \Psi[S]^{\otimes n} = \log \Psi[S] \otimes I \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes I \otimes \log \Psi[S],$$

and introducing the operator $F = \Psi^*[\log \Psi[S]]$, where Ψ^* is the dual channel, we can rewrite the second term as

$$\begin{aligned} n \text{Tr} \frac{(F \otimes I \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes I \otimes F)}{n} (\bar{S}^{n,\delta} - S^{\otimes n}) \\ = \frac{n}{d_{n,\delta}} \sum_{j \in \hat{T}^{n,\delta}} \left[\frac{f(j_1) + \cdots + f(j_n)}{n} - \sum_{j=1}^d \lambda_j f(j) \right], \quad (9.40) \end{aligned}$$

where $f(j) = \langle e_j | F | e_j \rangle$, which is evaluated by $n\delta d \max f$ via (9.35). This establishes (9.38) in the case of a nondegenerate $\Psi[S]$.

Coming back to the general case, let us denote by P_Ψ the supporting projector of $\Psi[S]$. In this case, the supporting projector of $\Psi[S]^{\otimes n}$ is $P_\Psi^{\otimes n}$, and the support of $\Psi^{\otimes n}(\bar{S}^{n,\delta})$ is contained in the support of $\Psi[S]^{\otimes n} = \Psi^{\otimes n}[S^{\otimes n}]$, because the support of $\bar{S}^{n,\delta}$ is contained in that of $S^{\otimes n}$. Thus, the second term in (9.39) can be understood as

$$\text{Tr} P_\Psi^{\otimes n} \log \left[P_\Psi^{\otimes n} \Psi[S]^{\otimes n} P_\Psi^{\otimes n} \right] P_\Psi^{\otimes n} \left(\Psi^{\otimes n} [\bar{S}^{n,\delta}] - \Psi[S]^{\otimes n} \right),$$

where now we have the logarithm of a nondegenerate operator in $P_\Psi^{\otimes n} \mathcal{H}_{A'}^{\otimes n}$. We can then repeat the argument, with F defined as $\Psi^*[P_\Psi(\log P_\Psi \Psi[S] P_\Psi) P_\Psi]$. This completes the proof of (9.34), from which (9.31) follows. \square

9.5 Notes and references

1. Theorem 9.1 was announced by Bennett, Shor, Smolin and Thapliyal in [24], and was proved in the paper [25] by the same authors.
2. Information capacities of quantum observables were studied by Dall'Arno, D'Ariano, Sacchi [46] and by Holevo [108]. The identity (9.15) was obtained in [46] and (9.16) was obtained in [108]. Relation (9.14) is due to Shirokov [186] who also obtained a general criterion for the coincidence $C_{ea} = C_\chi$, saying that the essential part of the channel should be the c-q channel [187]. The duality between quantum observables and ensembles was introduced by Hall [74]. The value (9.21) was obtained in the paper of Jozsa, Robb and Wootters [124] as the “subentropy” of the chaotic state $S = I/d$.
3. We here provide a simplified proof, following [101]. Concerning strongly typical projectors and the solution of Exercise 9.10, see [44].

Chapter 10

Transmission of quantum information

As was already stressed, a quantum state is itself a special kind of information resource, insofar as it contains statistical uncertainty. Statistical mechanics studies irreversible state changes of a physical system that result from its interaction with the environment (and accompanied by loss of information). Information theory in a sense solves a reverse problem, namely that of how to reduce these losses to a negligible quantity by using a special processing of the system state – encoding before and decoding after the irreversible evolution described by a noisy channel. This chapter is devoted to this kind of problems.

We shall start with a brief review of the methods of encoding the quantum information resilient to a certain specific kind of errors (e.g. an arbitrary error in one arbitrary qubit), when complete error correction is possible. We then pass to an approximate setting, where some measure of fidelity for the processing of a quantum state is introduced. In this situation, the aim is to asymptotically achieve an exact transmission of the quantum state through the composite channel $\Phi^{\otimes n}$ with maximal possible rate as $n \rightarrow \infty$. In this way, another analog of the Shannon Coding Theorem emerges, concerning the transmission of quantum information. This leads to the notion of the quantum capacity, expressed in terms of an entropic quantity called coherent information. On the other hand, the quantum capacity of a channel turns out to be closely related to its cryptographic characteristics, such as the capacity for the secret transmission of classical information (the private classical capacity). The role of the eavesdropper in such scenarios is played by the environment of the quantum system.

10.1 Quantum error-correcting codes

10.1.1 Error correction by repetition

When transmitting information, one would like to have a code that is robust against errors. In the classical case, the existence of such codes for rates below the capacity follows from the Shannon Coding Theorem. However, this theorem does not provide a constructive method for the design of such codes, and a practical solution to this problem is the subject of extensive research in classical information theory.

A straightforward method of reducing errors is repetition of messages, which, of course, reduces the rate of transmission. Take the binary alphabet $0, 1$ and assume that the probability p of a bit flip in the process of transmission is small, so that the probability of a change in more than in one bit is negligible. Consider the code

$0 \rightarrow 00, 1 \rightarrow 11$. In this case, having received 00 or 11, one can conclude with a high degree of certitude that the encoded symbol was 0 or, correspondingly, 1. However, if 01 or 10 was received, one can only ascertain an error but not recover the encoded symbol. However, this defect is easily remedied by adding one more bit to the code, $0 \rightarrow 000$ and $1 \rightarrow 111$. The last code corrects the arbitrary error in one arbitrary bit of the received message.

A direct generalization of this recipe to the quantum case is impossible, because quantum information cannot be copied. By the very nature of quantum information, one should be able to transmit robustly not only the basis states but also their superposition. Although at first glance this seems not feasible, the problem has an ingenious solution. Let us consider an example of solving such a problem in an important model case.

Assume that one wants to transmit an arbitrary pure state of a qubit $\psi = a|0\rangle + b|1\rangle$, provided that encoding by states of several qubits is allowed. The encoding in the quantum case is just an isometric mapping. Following the classical analogy, consider first the code that maps the basis vectors as follows

$$|0\rangle \rightarrow |000\rangle, \quad |1\rangle \rightarrow |111\rangle. \quad (10.1)$$

Such a code corrects the “bit flip”, i.e. the change $|0\rangle \leftrightarrow |1\rangle$ in any one of the qubits. We are interested in the arbitrary state $a|0\rangle + b|1\rangle$, which is now encoded into $a|000\rangle + b|111\rangle$. Let, for example, the bit flip occur in the first qubit:

$$a|0\rangle + b|1\rangle \rightarrow a|100\rangle + b|011\rangle,$$

In this case, the states $a|000\rangle + b|111\rangle, a|100\rangle + b|011\rangle$ are orthogonal, and hence can be perfectly distinguished to correct the error.

However, such a code does not correct a “phase flip” $|0\rangle \leftrightarrow |0\rangle, |1\rangle \leftrightarrow -|1\rangle$. Indeed, after such a phase error in one bit, we have $a|000\rangle - b|111\rangle$ instead of $a|000\rangle + b|111\rangle$, and in general these states are not orthogonal, and hence not perfectly distinguishable.

Now remark that the Hadamard transform

$$|0\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |1\rangle \rightarrow \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$$

maps a phase flip into a bit flip and vice versa. By converting the code (10.1), we obtain another code

$$\begin{aligned} |0\rangle &\rightarrow \frac{1}{2^{3/2}}(|0\rangle + |1\rangle)(|0\rangle + |1\rangle)(|0\rangle + |1\rangle), \\ |1\rangle &\rightarrow \frac{1}{2^{3/2}}(|0\rangle - |1\rangle)(|0\rangle - |1\rangle)(|0\rangle - |1\rangle), \end{aligned} \quad (10.2)$$

which corrects a phase flip in one arbitrary qubit, but does not correct bit flips.

The *Shor code*, which corrects both phase and bit flips in one qubit, is obtained by concatenation of the codes (10.1), (10.2), and requires 9 qubits

$$\begin{aligned} |0\rangle &\rightarrow \frac{1}{2^{3/2}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ |1\rangle &\rightarrow \frac{1}{2^{3/2}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle) \end{aligned} \quad (10.3)$$

Remarkably, it turns out that this code corrects not only bit and phase errors but also an *arbitrary* error in one of the nine qubits, since it can be represented as a combination of these two basic errors. This follows from the fact that the general conditions for error correction discussed in Section 10.1.3 are satisfied.

10.1.2 General formulation

Let S be an arbitrary state in a Hilbert space \mathcal{M} . A *quantum code* is an isometric map $V : \mathcal{M} \rightarrow \mathcal{N}$, transforming states S into encoded states VSV^* in another Hilbert space \mathcal{N} . The system \mathcal{N} is subject to errors, the effect of which is described by completely positive maps Φ in $\mathfrak{T}(\mathcal{N})$, that belong to a certain class \mathcal{E} . As we know, irreversible evolutions are described by CP maps that satisfy the normalization $\Phi(I) \leq I$ (see Section 6.5), but in the present situation this restriction is not essential, since the effect of an error is determined up to a constant factor.

Thus, transformations of the quantum information are given by the diagram

$$S \longrightarrow VSV^* \longrightarrow \Phi(VSV^*), \quad \Phi \in \mathcal{E}.$$

Definition 10.1. The code V is *correcting errors from* \mathcal{E} , if there exists a *reverse channel* \mathcal{D} such that

$$\mathcal{D}[\Phi(VSV^*)] = c(\Phi)S, \quad (10.4)$$

for an arbitrary state S and an arbitrary $\Phi \in \mathcal{E}$, where $c(\Phi)$ is a constant depending only on Φ .

In fact, the code is determined by the subspace $\mathcal{L} = V\mathcal{M} \subset \mathcal{N}$ and can be defined as such, without introducing \mathcal{M} and V .

Exercise 10.2. *Storage of quantum information in the memory of a quantum computer.* Let $\mathcal{N} = \mathcal{H}_2^{\otimes n}$ be the quantum register in which the quantum information from the subspace \mathcal{M} is to be stored. Consider errors involving no more than m qubits of the register. The corresponding class $\mathcal{E}(n, m)$ consists of the maps $\Phi = \Phi_1 \otimes \cdots \otimes \Phi_n$, where the number of $\Phi_k \neq \text{Id}$ is not greater than m , while the error Φ_k in the k -th qubit can be an arbitrary completely positive map. Operators of elementary errors in each qubit are usually given by the Pauli

matrices, with σ_x describing the bit flip, σ_z the phase flip, and $\sigma_y = i\sigma_x\sigma_z$ their combination. Together with the unit operator $\sigma_0 = I$, corresponding to absence of error, they form a basis in the algebra of observables of the qubit.

The Shor code demonstrates the possibility of error-correction for the class $\mathcal{E}(n, 1)$, if n is large enough (one can show that the smallest possible value of n for a code that corrects one error is 5). Restriction to the correction of only one error is, of course, essential here. However, it can be shown that there exist codes that correct errors from $\mathcal{E}(n, m)$, where the ratio m/n is greater than a certain positive number for large enough n .

10.1.3 Necessary and sufficient conditions for error correction

The class \mathcal{E} of errors usually has the following structure. It consists of mappings of the form $\Phi(S) = \sum_j V_j S V_j^*$, where $V_j \in \text{Lin}(B_1, \dots, B_p)$, and B_j are the fixed, linearly independent operators of elementary errors.

Theorem 10.3 (Knill and Laflamme [133]). *The following statements are equivalent:*

- i. the code \mathcal{L} corrects errors from \mathcal{E} ;
- ii. the code \mathcal{L} corrects the error $\Phi[S] = \sum_{j=1}^p B_j S B_j^*$;
- iii. for all $\phi, \psi \in \mathcal{L}$ such that $\langle \phi | \psi \rangle = 0$, one has $\langle \phi | B_i^* B_j | \psi \rangle = 0$ for all $i, j = 1, \dots, p$.
- iv. for an orthonormal basis $\{|k\rangle\}$ in \mathcal{L} one has

$$\begin{aligned} \langle k | B_i^* B_j | l \rangle &= \langle l | B_i^* B_j | l \rangle, \quad \text{for all } i, j; k, l, \\ \langle k | B_i^* B_j | l \rangle &= 0, \quad \text{for } k \neq l; \end{aligned}$$

- v. $P_{\mathcal{L}} B_i^* B_j P_{\mathcal{L}} = b_{ij} P_{\mathcal{L}}$, where $P_{\mathcal{L}}$ is the projector onto \mathcal{L} .

Proof.

i. \Rightarrow ii. is obvious.

ii. \Rightarrow iii. Let there exist a reverse channel $\mathcal{D}[S] = \sum_r R_r S R_r^*$ for the error Φ . Consider a pure input state $S = |\psi\rangle\langle\psi|$, $|\psi\rangle \in \mathcal{L}$. In this case,

$$\sum_r \sum_j R_r B_j |\psi\rangle\langle\psi| B_j^* R_r^* = c |\psi\rangle\langle\psi|.$$

But this is possible only if $R_r B_j |\psi\rangle = c_{jr} |\psi\rangle$, in which case $c = \sum_r \sum_j |c_{jr}|^2$.

Let $|\phi\rangle, |\psi\rangle \in \mathcal{L}$ be two orthogonal vectors. Since the reverse channel is trace preserving, we have $I = \sum_r R_r^* R_r$. Therefore,

$$\langle \phi | B_i^* B_j | \psi \rangle = \sum_r \langle \phi | B_i^* R_r^* R_r B_j | \psi \rangle = \left(\sum_r \bar{c}_{ir} c_{jr} \right) \langle \phi | \psi \rangle = 0.$$

iii. \Rightarrow iv. The second equality follows trivially, while the first one is obtained by letting $|\phi\rangle = |k + l\rangle$, $|\psi\rangle = |k - l\rangle$.

iv. \Rightarrow v. Denoting $b_{ij} = \langle k | B_i^* B_j | k \rangle$, we can write the condition iv. as

$$\langle k | B_i^* B_j | l \rangle = \delta_{kl} b_{ij},$$

which is equivalent to v.

v. \Rightarrow i. The matrix $[b_{ij}]$ is Hermitian positive, hence there exists a unitary matrix $[u_{ij}]$ such that

$$\sum_{ij} \bar{u}_{ir} b_{ij} u_{js} = \delta_{rs} \lambda_r,$$

where $\lambda_r \geq 0$. Introducing $\tilde{B}_s = \sum_j u_{js} B_j$, we have $\Phi[S] = \sum_{s=1}^p \tilde{B}_s S \tilde{B}_s^*$ and

$$P_{\mathcal{L}} \tilde{B}_r^* \tilde{B}_s P_{\mathcal{L}} = \delta_{rs} \lambda_r P_{\mathcal{L}}.$$

Thus, for $\lambda_r > 0$ we have

$$\lambda_s^{-1/2} \tilde{B}_s P_{\mathcal{L}} = U_s P_{\mathcal{L}},$$

where U_s are the partially isometric operators that map \mathcal{L} onto mutually orthogonal subspaces $\mathcal{L}_s \subset \mathcal{H}$. It follows that

$$\Phi[S] = \sum_{s=1}^p \lambda_s U_s S U_s^* \quad (10.5)$$

for any state S with $\text{supp } S \subset \mathcal{L}$.

Let P_s be the projectors onto the subspaces \mathcal{L}_s , and denote $P_0 = I - \sum_s P_s$. We define the channel

$$\mathcal{D}[S] = \sum_s U_s^* P_s S P_s U_s + P_0 S P_0$$

and wish to show that it is the reverse for all errors in \mathcal{E} . However, a linear combination of elementary errors B_j is also a linear combination of \tilde{B}_r and, taking into account that $U_s^* P_s \tilde{B}_r P_{\mathcal{L}} = \lambda_r^{1/2} \delta_{sr} P_{\mathcal{L}}$, we obtain (10.4) for arbitrary $\Phi \in \mathcal{E}$. \square

Exercise 10.4. Check the conditions iv. for the Shor code (10.3) and the elementary error operators given by the Pauli matrices in arbitrary one qubit.

10.1.4 Coherent information and perfect error correction

We now consider an arbitrary channel Φ from A to B . Let $S = S_A$ be an input state of the channel. Introduce the reference system R so that $S_{AR} = |\psi_{AR}\rangle\langle\psi_{AR}|$ is a purification of S_A . Let $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ be the Stinespring isometry (6.7) for the channel Φ , where E is the environment.

An important component of the mutual information $I(S, \Phi)$ is the *coherent information*

$$\begin{aligned} I_c(S, \Phi) &= H(\Phi[S]) - H(S; \Phi) \\ &= H(B) - H(E) \\ &= H(B) - H(RB) \\ &= -H(R|B). \end{aligned}$$

As we shall see later, it is closely related to the quantum capacity of the channel Φ .

Somewhat disappointingly, the coherent information *does not* have several “natural” information properties, such as concavity in S , subadditivity, or the second chain rule. Moreover, it can take negative values. Its classical analog is never positive, $H(Y) - H(XY) = -H(X|Y) \leq 0$. However, $I_c(S, \Phi)$ is convex in Φ and satisfies the first chain rule:

$$I_c(S, \Phi_2 \circ \Phi_1) \leq I_c(S, \Phi_1), \quad (10.6)$$

as can be seen from the relation $I_c(S, \Phi) = I(S, \Phi) - H(S)$, and from the corresponding properties for the quantum mutual information $I(S, \Phi)$. Let us see why one should not expect the second chain rule: $I_c(S, \Phi_2 \circ \Phi_1) \leq I_c(\Phi_1[S], \Phi_2)$. Assuming that this holds, we would have

$$H(\Phi_2 \circ \Phi_1[S]) - H(S, \Phi_2 \circ \Phi_1) \leq H(\Phi_2 \circ \Phi_1[S]) - H(\Phi_1[S], \Phi_2),$$

i.e. $H(S, \Phi_2 \circ \Phi_1) \geq H(\Phi_1[S], \Phi_2)$. But this is the same as $H(E_1 E_2) \geq H(E_2)$, where E_j is the environment for j -th channel. However, such a monotonicity does not hold in general for quantum entropy, as we have noticed in Section 7.5.

Exercise 10.5. Show that subadditivity of $I_c(S, \Phi)$ is equivalent to the following inequality

$$H(B_1 B_2) - H(B_1) - H(B_2) \leq H(E_1 E_2) - H(E_1) - H(E_2),$$

which need not hold in general.

There is a close relationship between the perfect transmission of quantum information, error-correction and the coherent information.

Definition 10.6. The channel Φ is called *perfectly reversible on the state* $S = S_A$, if there is a *reverse channel* \mathcal{D} from B to A , such that

$$(\mathcal{D} \circ \Phi \otimes \text{Id}_R)[S_{AR}] = S_{AR}.$$

Proposition 10.7. *The following conditions are equivalent:*

- i. *the channel Φ is perfectly reversible on the state S*
- ii. $I(R; E) = 0$, i.e. $S_{RE} = S_R \otimes S_E$ (where $S_{RE} = \text{Tr}_B S_{BRE}$)
- iii. $I_c(S, \Phi) = H(S)$.
- iv. *the channel Φ admits a representation (10.5) on the support of the state S*

The meaning of condition ii. is that no information goes to the environment, i.e. the channel is “private”. Thus, the perfect reversibility of the channel and its perfect privacy are equivalent. Condition iii. means that for perfect transmission by the channel Φ of the state S , the coherent information $I_c(S, \Phi)$ should reach its maximal possible value $H(S)$. In particular, by taking $S = \bar{S}_A = I_A/d_A$ – the chaotic state, we obtain the condition $\log d_A = I_c(\bar{S}_A, \Phi)$. This indicates that the coherent information is the quantity relevant to the quantum capacity of the channel Φ , i.e. the capacity for perfect transmission of quantum states. In the next sections, we shall see that for approximately perfect transmission by the memoryless channel $\Phi^{\otimes n}$, an appropriate analog of these conditions should hold asymptotically, with $n \rightarrow \infty$.

Proof.

i. \Rightarrow ii. The vector $|\psi_{BRE}\rangle = (V \otimes I_R)|\psi_{AR}\rangle$ is a vector of the pure state S_{BRE} of the composite system BRE . The perfect reversibility implies

$$(\mathcal{D} \otimes \text{Id}_{RE})[S_{BRE}] = S_{ARE}.$$

Since the partial state S_{AR} is pure, by Corollary 3.12 $S_{ARE} = S_{AR} \otimes S_E$. Taking the partial trace with respect to A , we obtain

$$S_{RE} = S_R \otimes S_E, \tag{10.7}$$

that is ii.

ii. \Leftrightarrow iii. We have

$$H(S) - I_c(S, \Phi) = H(A) + H(E) - H(B) \tag{10.8}$$

$$= H(R) + H(E) - H(RE) = I(R; E) \geq 0; \tag{10.9}$$

with equality if and only if $I(R; E) = 0$, i.e. $S_{RE} = S_R \otimes S_E$.

ii. \Rightarrow i. The vector $|\psi_{BRE}\rangle = (V \otimes I_R)|\psi_{AR}\rangle$ is the vector of joint pure state S_{BRE} of the system BRE , which is a purification of the state S_{RE} . Starting from the right hand side of (10.7), we obtain another purification $|\psi_{AR}\rangle \otimes |\psi_{EE'}\rangle$, where E' is a purifying system for S_E . By choice of E' we can always assume that this second purification has dimensionality not less than that of the first. Hence, by the remark after Theorem 3.10, there is an isometric operator $W : \mathcal{H}_B \rightarrow \mathcal{H}_A \otimes \mathcal{H}_{E'}$ such that

$$(I_{RE} \otimes W)|\psi_{BRE}\rangle = |\psi_{AR}\rangle \otimes |\psi_{EE'}\rangle. \quad (10.10)$$

Taking the partial trace with respect to E, E' of the corresponding density operator, we obtain the perfect reversibility

$$(\text{Id}_R \otimes \mathcal{D} \circ \Phi)[S_{AR}] = S_{AR}, \quad (10.11)$$

where the reverse channel is defined as

$$\mathcal{D}[S_B] = \text{Tr}_{E'} W S_B W^*. \quad (10.12)$$

i. \Leftrightarrow iv. Later, we shall prove Corollary 10.14, the statement i. \Leftrightarrow iii. of which, with the replacement of Φ by $\mathcal{D} \circ \Phi$, implies that the channel Φ is perfectly reversible on the state S if and only if

$$\mathcal{D} \circ \Phi[\tilde{S}] = \tilde{S}$$

for all \tilde{S} , with $\text{supp } \tilde{S} \subset \mathcal{L} \equiv \text{supp } S$, where $\text{supp } S$ is the support of the density operator S (see Section 1.3). We can reformulate this property by saying that Φ is *perfectly reversible on the subspace \mathcal{L}* . In other words, the subspace \mathcal{L} is a quantum code that corrects the error Φ (and hence, all errors of the associated class \mathcal{E}). The statement then follows from the proof of Theorem 10.3. \square

The following proposition provides yet another characterization in terms of complementary channels (see Section 6.6), which prepares the ground for the Coding Theorem for the private classical capacity of a quantum channel, in Section 10.4.1.

Proposition 10.8. *For a given Stinespring dilation of the channel Φ the following conditions are equivalent,*

- i. *the channel Φ is perfectly reversible on the subspace \mathcal{L}*
- ii. *the complementary channel $\tilde{\Phi}$ is completely depolarizing on \mathcal{L} , i.e.*

$$\tilde{\Phi}[\tilde{S}] = S_E, \quad (10.13)$$

for any state \tilde{S} with $\text{supp } \tilde{S} \subset \mathcal{L}$, where S_E is a fixed state.

Proof.

i. \Rightarrow ii. Assuming Φ is perfectly reversible, the relation (10.10) holds, with an isometric W for all $|\psi_{RA}\rangle$ having a tensor decomposition with A -components in \mathcal{L} . Here, $|\psi_{EE'}\rangle$ cannot depend on the input $|\psi_{RA}\rangle$, since otherwise the right-hand side would be nonlinear in $|\psi_{RA}\rangle$. Taking the partial trace with respect to RAE' of the corresponding density operators, we obtain (10.13).

ii. \Rightarrow i. We have a Stinespring dilation $\tilde{\Phi}[\tilde{S}] = \text{Tr}_B V \tilde{S} V^*$, where B plays the role of an “environment” for the environment system. Let (10.13) hold and let $|\psi_{EE'}\rangle$ be a purification of S_E . In this case, there is another Stinespring dilation for the restriction of the channel $\tilde{\Phi}$ to the states with support in \mathcal{L} :

$$\tilde{\Phi}[\tilde{S}] = \text{Tr}_{AE'} V' \tilde{S} V'^*, \quad (10.14)$$

where $V' = I_A \otimes |\psi_{EE'}\rangle$ is an isometric operator from \mathcal{H}_A to $\mathcal{H}_{AAE'} = \mathcal{H}_E \otimes \mathcal{H}_{AE'}$, acting as $V'|\psi\rangle = |\psi\rangle \otimes |\psi_{EE'}\rangle$. In this dilation, the role of an “environment” for the environment E is now played by AE' . By Theorem 6.12, there is a partial isometry $W : \mathcal{H}_B \rightarrow \mathcal{H}_{AE'}$ such that

$$(W \otimes I_E)V = V' = I_A \otimes |\psi_{EE'}\rangle.$$

It follows that relation (10.10) holds for all $|\psi_{RA}\rangle$ having a tensor decomposition with A -components in \mathcal{L} . By possibly extending the space $\mathcal{H}_{E'}$ we can replace W with an isometry that preserves relation (10.10). Following a similar argument as in the proof of the previous proposition, we take the partial trace of the corresponding density operator with respect to REE' and obtain the perfect reversibility on \mathcal{L} :

$$\mathcal{D} \circ \Phi[\tilde{S}] = \tilde{S}, \quad (10.15)$$

where the reverse channel is defined by (10.12). \square

10.2 Fidelities for quantum information

As a measure of exactness of the transmission of a quantum state S by the channel Φ one could take the trace norm $\|S - \Phi[S]\|_1$. However, in quantum information theory there are a number of other useful quantities playing the same role as the probability of a correct decision in the classical case. Let us first consider these quantities and study the relations between them.

10.2.1 Fidelities for pure states

Lemma 10.9. *Let ψ be a unit vector and S an arbitrary state. In this case, the following inequalities hold*

$$2[1 - \langle \psi | S | \psi \rangle] \leq \|\psi\rangle\langle\psi| - S\|_1 \leq 2\sqrt{1 - \langle \psi | S | \psi \rangle}. \quad (10.16)$$

If $S = |\phi\rangle\langle\phi|$ is a pure state, the second inequality turns into equality, i.e.

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_1 = 2\sqrt{1 - |\langle\psi|\phi\rangle|^2}. \quad (10.17)$$

Proof. By using (1.19) we have

$$\| |\psi\rangle\langle\psi| - S \|_1 = \max_U |\text{Tr}(|\psi\rangle\langle\psi| - S)U|,$$

where the maximum is over all unitary operators U . Taking $U = 2|\psi\rangle\langle\psi| - I$ produces the first inequality.

The equality (10.17) is obtained similar to relation (2.44). To compute the trace norm, it is sufficient to find the eigenvalues of the operator $|\psi\rangle\langle\psi| - |\phi\rangle\langle\phi|$, which has rank 2 (see Proposition 2.26).

Now, let S be an arbitrary density operator and consider its spectral decomposition $S = \sum \lambda_i S_i$. By using the convexity of the norm, the concavity of the square root, and relation (10.17), we obtain

$$\begin{aligned} \| |\psi\rangle\langle\psi| - S \|_1 &\leq \sum_i \lambda_i \| |\psi\rangle\langle\psi| - S_i \|_1 \\ &= 2 \sum_i \lambda_i \sqrt{1 - \langle\psi|S_i|\psi\rangle} \leq 2\sqrt{1 - \langle\psi|S|\psi\rangle}. \end{aligned}$$

□

For a pure state $|\psi\rangle\langle\psi|$ and an arbitrary state S , the quantity

$$F(|\psi\rangle\langle\psi|, S) = \langle\psi|S|\psi\rangle \quad (10.18)$$

is called the *fidelity* between the pure state $|\psi\rangle\langle\psi|$ and the state S . Clearly, $F \leq 1$, with equality if and only if $S = |\psi\rangle\langle\psi|$. Note that in quantum data compression (Section 5.5) we in fact used the fidelity (10.18). There, the state $|\psi_i\rangle\langle\psi_i|$ appeared with probability p_i . Hence, the average fidelity is just $\bar{F} = \sum_i p_i \langle\psi_i|S_i|\psi_i\rangle$.

In connection with error-correcting codes, a natural definition of fidelity is as follows. Let a subspace (quantum code) $\mathcal{L} \subset \mathcal{H}$ and a channel Φ in \mathcal{H} , for which the input and output spaces coincide $\mathcal{H}_B = \mathcal{H}_A = \mathcal{H}$, be given. The *subspace fidelity* is defined as

$$\begin{aligned} F_s(\mathcal{L}, \Phi) &= \min_{\psi \in \mathcal{L}, \|\psi\|=1} F(|\psi\rangle\langle\psi|, \Phi[|\psi\rangle\langle\psi|]) \\ &= \min_{\psi \in \mathcal{L}, \|\psi\|=1} \langle\psi| \Phi[|\psi\rangle\langle\psi|] |\psi\rangle. \end{aligned}$$

By Lemma 10.9,

$$2[1 - F_s(\mathcal{L}, \Phi)] \leq \max_{\psi \in \mathcal{L}, \|\psi\|=1} \| |\psi\rangle\langle\psi| - \Phi[|\psi\rangle\langle\psi|] \|_1 \leq 2\sqrt{1 - F_s(\mathcal{L}, \Phi)}. \quad (10.19)$$

Another important quantity is the *entanglement fidelity* $F_e(S, \Phi)$, defined as follows. Let us purify the state $S = S_A$ to $|\psi_{AR}\rangle\langle\psi_{AR}|$ in the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_R$ and consider the fidelity for this pure state under the action of the trivial extension of the channel

$$F_e(S, \Phi) = \langle\psi_{AR}|(\Phi \otimes \text{Id}_R)(|\psi_{AR}\rangle\langle\psi_{AR}|)|\psi_{AR}\rangle.$$

There is another convenient expression for $F_e(S, \Phi)$, which also shows independence of this definition of the way of purification of the initial state.

Lemma 10.10. *Let the channel have the Kraus representation*

$$\Phi[S] = \sum_i V_i S V_i^*, \quad (10.20)$$

then

$$F_e(S, \Phi) = \sum_i |\text{Tr } V_i S|^2. \quad (10.21)$$

Proof. Indeed,

$$\begin{aligned} \langle\psi_{AR}| \sum_i (V_i \otimes I_R) |\psi_{AR}\rangle\langle\psi_{AR}| (V_i \otimes I_R)^* |\psi_{AR}\rangle &= \sum_i |\langle\psi_{AR}| (V_i \otimes I_R) |\psi_{AR}\rangle|^2 \\ &= \sum_i |\text{Tr } V_i S|^2. \quad \square \end{aligned}$$

Notice that for a pure state $S = |\psi\rangle\langle\psi|$,

$$F_e(S, \Phi) = \langle\psi|\Phi[|\psi\rangle\langle\psi|]|\psi\rangle = F(|\psi\rangle\langle\psi|, \Phi[|\psi\rangle\langle\psi|]).$$

Exercise 10.11. Prove that $F_e(S, \Phi)$ is a convex function of S . *Hint:* By (10.21), $F_e(S, \Phi)$ is the sum of squares of affine functions of S .

10.2.2 Relations between the fidelity measures

Lemma 10.12. *For an arbitrary state S*

$$1 - F_e(S, \Phi) \leq 4\sqrt{1 - F_s(\text{supp } S, \Phi)}.$$

Proof. We have

$$\begin{aligned} 1 - F_e(S, \Phi) &= 1 - \langle\psi_{AR}|(\Phi \otimes \text{Id}_R)(|\psi_{AR}\rangle\langle\psi_{AR}|)|\psi_{AR}\rangle \\ &= \langle\psi_{AR}|((\text{Id}_A - \Phi) \otimes \text{Id}_R)[|\psi_{AR}\rangle\langle\psi_{AR}|]| \psi_{AR}\rangle. \end{aligned}$$

Representing $|\psi_{AR}\rangle = \sum_j |\psi_j\rangle \otimes |e_j\rangle$, where $\{|e_j\rangle\}$ is an orthonormal basis in \mathcal{H}_R , $\sum_j \|\psi_j\|^2 = 1$, $\psi_j \in \text{supp } S$, we obtain the equivalent relation

$$\sum_{jk} \langle \psi_j | (\text{Id} - \Phi) [|\psi_j\rangle \langle \psi_k|] | \psi_k \rangle = \sum_{jk} \text{Tr} |\psi_k\rangle \langle \psi_j| (|\psi_j\rangle \langle \psi_k| - \Phi [|\psi_j\rangle \langle \psi_k|]).$$

Taking into account inequality (1.18) and the fact that the operators $|\psi_j\rangle \langle \psi_k|$ all have both trace and operator norms equal to $\|\psi_j\| \|\psi_k\|$, we obtain that this does not exceed

$$\max_T \frac{\|T - \Phi[T]\|_1}{\|T\|_1},$$

where the maximum is taken over all nonzero operators T acting in $\mathcal{L} = \text{supp } S$. By decomposing $T = T_1 + iT_2$, where $T_1^* = T_1$, $T_2^* = T_2$ again act in \mathcal{L} , and taking into account that $\|T_{1,2}\|_1 \leq \|T\|_1$, as well as the triangle inequality, we infer that this does not exceed the following quantity

$$2 \max_{T^*=T} \frac{\|T - \Phi[T]\|_1}{\|T\|_1} = 2 \max_{S: \text{supp } S \in \mathcal{L}} \|S - \Phi[S]\|_1. \quad (10.22)$$

Here, the inequality \geq is obtained by restricting ourselves to positive T . On the other hand,

$$\max_{T^*=T} \frac{\|T - \Phi[T]\|_1}{\|T\|_1} \leq \|S_+ - \Phi[S_+]\|_1 \frac{p_+}{p_+ + p_-} + \|S_- - \Phi[S_-]\|_1 \frac{p_-}{p_+ + p_-}, \quad (10.23)$$

where we denoted $p_{\pm} = \text{Tr } T_{\pm}$; $S_{\pm} = p_{\pm}^{-1} T_{\pm}$. Here, the right hand side is less than or equal to the right hand side in the relation (10.22), which is thus proved. Finally,

$$\max_{S: \text{supp } S \in \mathcal{L}} \|S - \Phi[S]\|_1 = \max_{\psi: \psi \in \mathcal{L}, \|\psi\|=1} \| |\psi\rangle \langle \psi| - \Phi[|\psi\rangle \langle \psi|] \|_1,$$

since the maximum of a convex function is attained at an extreme point of the convex set. By using the second inequality from (10.19), we obtain that the expression does not exceed $4\sqrt{1 - F_s(\mathcal{L}, \Phi)}$. \square

Exercise 10.13. By using the convexity property from Exercise 10.11, prove the inequality

$$1 - F_s(\mathcal{L}, \Phi) \leq 1 - \min_{S: \text{supp } S \subset \mathcal{L}} F_e(S, \Phi). \quad (10.24)$$

Now, let $\mathcal{L} = \mathcal{H}$. In this case,

$$\max_T \frac{\|T - \Phi[T]\|_1}{\|T\|_1} = \|\text{Id} - \Phi\|.$$

From the proof of Lemma 10.12, we have the chain of inequalities

$$1 - \min_S F_e(S, \Phi) \leq \|\text{Id} - \Phi\| \leq 2 \max_{\|\psi\|=1} \| |\psi\rangle\langle\psi| - \Phi[|\psi\rangle\langle\psi|] \|_1 \leq 2\|\text{Id} - \Phi\|. \quad (10.25)$$

Together with relation (10.24) this means that a deviation of the channel Φ from the ideal channel Id is equivalently described by one of the quantities

$$\|\text{Id} - \Phi\|; \quad \max_S \|S - \Phi[S]\|_1; \quad 1 - F_S(\mathcal{H}, \Phi); \quad 1 - \min_S F_e(S, \Phi).$$

In the consideration of the quantum capacity we will need the following obvious corollary of Lemma 10.12:

$$1 - F_e(\bar{S}, \Phi) \leq 4\sqrt{1 - F_S(\mathcal{H}, \Phi)}, \quad (10.26)$$

where \bar{S} is the chaotic state in \mathcal{H} .

It is instructive to consider the “ideal” case of the unit fidelities.

Corollary 10.14. *The following conditions are equivalent,*

- i. $F_e(S, \Phi) = 1$
- ii. $F_S(\text{supp } S, \Phi) = 1$
- iii. $\Phi[\tilde{S}] = \tilde{S}$ for arbitrary state \tilde{S} with $\text{supp } \tilde{S} \subset \text{supp } S$
- iv. $F_e(\tilde{S}, \Phi) = 1$ for arbitrary state \tilde{S} with $\text{supp } \tilde{S} \subset \text{supp } S$

Proof. Let us prove i. \Rightarrow ii. Let $|\psi\rangle \in \text{supp } S$. In this case, we can write

$$S = p|\psi\rangle\langle\psi| + (1 - p)S',$$

where $p > 0$ and S' is a density operator. By Exercise 10.11, the function $S \rightarrow F_e(S, \Phi)$ is convex. Therefore,

$$pF_e(|\psi\rangle\langle\psi|, \Phi) + (1 - p)F_e(S', \Phi) \geq F_e(S, \Phi) = 1,$$

and hence $F_e(|\psi\rangle\langle\psi|, \Phi) = 1$. But this means that

$$\langle\psi| \Phi[|\psi\rangle\langle\psi|] |\psi\rangle = 1$$

for all $|\psi\rangle \in \text{supp } S$ and $F_S(\text{supp } S, \Phi) = 1$.

Lemma 10.12 implies ii. \Rightarrow i. Moreover, since ii. obviously implies the same condition, with $\text{supp } \tilde{S}$ instead of $\text{supp } S$, it also implies iv., which is formally stronger than i.

By the definition of F_S , ii. is equivalent to iii. for pure, and, hence, for mixed states \tilde{S} . □

10.2.3 Fidelity and the Bures distance

Here, we discuss the fidelity between two arbitrary mixed states. This material will be needed only in Section 10.4.4.

For pure states $S = |\phi\rangle\langle\phi|$, $T = |\psi\rangle\langle\psi|$, the fidelity is equal to $F(T, S) = |\langle\phi|\psi\rangle|^2$, and formula (10.17) provides the relation between the fidelity and the trace norm distance between S, T . Now, let S, T be arbitrary density operators.

Definition 10.15. The *fidelity* between the density operators S, T is defined by the relation

$$F(T, S) = \max |\langle\psi_T|\psi_S\rangle|^2, \quad (10.27)$$

where the maximum is taken over all possible purifications ψ_S, ψ_T of the states S, T (in the same Hilbert space $\mathcal{H} \otimes \mathcal{H}'$.)

By using Theorem 3.11, we see that

$$F(T, S) = \max_W |\langle\psi_T|(I \otimes W)\psi_S\rangle|^2, \quad (10.28)$$

where now ψ_S, ψ_T are fixed purifications of the states S, T in a fixed Hilbert space $\mathcal{H} \otimes \mathcal{H}'$, and the maximum is taken over all unitary operators W in \mathcal{H}' . In particular, taking the purifications \sqrt{S}, \sqrt{T} in $L^2(\mathcal{H}) \simeq \mathcal{H} \otimes \mathcal{H}^*$ we obtain

$$F(T, S) = \max_W \left| \text{Tr } \sqrt{S} \sqrt{T} W \right|^2,$$

where the maximum is over all unitary operators W in \mathcal{H} . By using the polar decomposition of the operator $\sqrt{S} \sqrt{T} = U |\sqrt{S} \sqrt{T}|$, one sees that the maximum is achieved for $W = U^*$ and is equal to

$$F(T, S) = \left(\text{Tr } |\sqrt{S} \sqrt{T}| \right)^2 = \left\| \sqrt{S} \sqrt{T} \right\|_1^2.$$

This coincides with (10.18) for $T = |\psi\rangle\langle\psi|$. Note that the above argument implies that in definition (10.27) we can restrict ourselves to maximization with respect to all purifications of only one of the states, with a fixed purification of another.

Exercise 10.16. Show that $F(T, S) \leq 1$, with equality if and only if $T = S$.

Related to the fidelity (10.27), there is a metric on the set of all states, called the *Bures distance*, namely

$$\beta(T, S) = \min \|\psi_T - \psi_S\|, \quad (10.29)$$

where the minimum is again taken over all possible purifications ψ_S, ψ_T of the states S, T . Taking the square of the norm and using the definition (10.27), one obtains

$$\beta(T, S) = \sqrt{2 \left(1 - \sqrt{F(T, S)} \right)} = \sqrt{2 \left(1 - \left\| \sqrt{S} \sqrt{T} \right\|_1 \right)}. \quad (10.30)$$

In particular, for pure states $T = |\psi\rangle\langle\psi|$, $S = |\phi\rangle\langle\phi|$

$$\beta(T, S) = \sqrt{2(1 - |\langle\psi|\phi\rangle|)}.$$

It also follows that, just as in the definition of the fidelity, in definition (10.29) we can restrict ourselves to minimization with respect to all purifications of only one of the states, with a fixed purification of another.

The triangle inequality for $\beta(T, S)$ then follows from the definition (10.29) and the corresponding inequality for the norm. For pure states, it reduces to

$$\sqrt{1 - |\langle\varphi|\psi\rangle|} \leq \sqrt{1 - |\langle\varphi|\chi\rangle|} + \sqrt{1 - |\langle\chi|\psi\rangle|}, \quad (10.31)$$

where φ, χ, ψ are arbitrary unit vectors.

Lemma 10.17. *For any two density operators S_1, S_2 in \mathcal{H}*

$$\beta(S_1, S_2)^2 \leq \|S_1 - S_2\|_1 \leq 2\beta(S_1, S_2), \quad (10.32)$$

The first inequality, combined with (10.30) and (10.28), implies

Corollary 10.18. *For any two density operators S_1, S_2 in \mathcal{H} and their given purifications ψ_{S_1}, ψ_{S_2} in $\mathcal{H} \otimes \mathcal{H}'$,*

$$\|S_1 - S_2\|_1 \geq 2 \left(1 - \max_W |\langle\psi_{S_1}|(I \otimes W)\psi_{S_2}\rangle| \right), \quad (10.33)$$

where the maximum is taken over all unitary operators W in \mathcal{H}' .

Proof of Lemma 10.17. Since $\text{Tr} \sqrt{S_1} \sqrt{S_2} \leq \text{Tr} |\sqrt{S_1} \sqrt{S_2}|$, to prove the first inequality in (10.32) it is sufficient to prove that

$$2 \left(1 - \text{Tr} \sqrt{S_1} \sqrt{S_2} \right) \leq \|S_1 - S_2\|_1.$$

This in turn follows from the more general inequality

$$\text{Tr} \left(\sqrt{S_1} - \sqrt{S_2} \right)^2 \leq \|S_1 - S_2\|_1,$$

valid for arbitrary positive S_1, S_2 . To prove this, for a Hermitian operator K , we denote by $1_+(K)$ (resp. $1_-(K)$) the spectral projector corresponding to positive (resp. non-positive) eigenvalues. In this case, $\sigma(K) = 1_+(K) - 1_-(K)$ is a unitary Hermitian operator that satisfies $\sigma(K)K = K\sigma(K) = |K|$. Denoting $K = \sqrt{S_1} - \sqrt{S_2}$, $L = \sqrt{S_1} + \sqrt{S_2}$, we have $S_1 - S_2 = \frac{1}{2}(KL + LK) \equiv K \circ L$. Hence,

$$\begin{aligned} \|S_1 - S_2\|_1 &\geq |\text{Tr} \sigma(K)K \circ L| = \text{Tr} |K|L \\ &= \text{Tr} |K| (1_+(K)L1_+(K) + 1_-(K)L1_-(K)). \end{aligned}$$

Since $L \geq K \geq -L$, then $1_+(K)L1_+(K) \geq 1_+(K)K$ and $1_-(K)L1_-(K) \geq -1_-(K)K$. Thus, $(1_+(K)L1_+(K) + 1_-(K)L1_-(K)) \geq |K|$, and therefore

$$\|S_1 - S_2\|_1 \geq \text{Tr } |K|^2 = \text{Tr } \left(\sqrt{S_1} - \sqrt{S_2} \right)^2.$$

To prove the second inequality in (10.32), let $S_1 = T_1^* T_1$, $S_2 = T_2^* T_2$ for some not necessarily Hermitian T_1, T_2 . Denoting $K = T_1 - T_2$, $L = T_1 + T_2$, we have

$$\begin{aligned} \|S_1 - S_2\|_1 &= \text{Tr } \sigma(S_1 - S_2)(S_1 - S_2) \\ &= \frac{1}{2} [\text{Tr } \sigma(S_1 - S_2)K^* L + \text{Tr } \sigma(S_1 - S_2)L^* K], \end{aligned}$$

By using the Cauchy-Schwarz inequality for the trace, we obtain

$$|\text{Tr } \sigma(S_1 - S_2)K^* L| \leq [\text{Tr } \sigma(S_1 - S_2)^2 K^* K \cdot \text{Tr } L^* L]^{\frac{1}{2}} = [\text{Tr } K^* K \cdot \text{Tr } L^* L]^{\frac{1}{2}};$$

and similarly for the second term. Thus, $\|S_1 - S_2\|_1 \leq [\text{Tr } K^* K \cdot \text{Tr } L^* L]^{\frac{1}{2}}$. Using the normalization of S_1, S_2 , we obtain

$$\text{Tr } K^* K = 2(1 - \Re \text{Tr } T_1^* T_2), \quad \text{Tr } L^* L = 2(1 + \Re \text{Tr } T_1^* T_2).$$

Take $T_1 = \sqrt{S_1}$, $T_2 = U \sqrt{S_2}$, where U is the unitary from the polar decomposition of $\sqrt{S_1} \sqrt{S_2}$, then $\Re \text{Tr } T_1^* T_2 = \Re \text{Tr } T_2^* T_1 = \text{Tr } |\sqrt{S_1} \sqrt{S_2}|$. Thus,

$$\begin{aligned} \text{Tr } K^* K &= 2 \left(1 - \left\| \sqrt{S_1} \sqrt{S_2} \right\|_1 \right) = \beta(S_1, S_2)^2, \\ \text{Tr } L^* L &= 2 \left(1 + \left\| \sqrt{S_1} \sqrt{S_2} \right\|_1 \right) \leq 4, \end{aligned}$$

and the second inequality in (10.32) follows. \square

10.3 The quantum capacity

10.3.1 Achievable rates

Let $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ be a quantum channel. Consider the composite channel $\Phi^{\otimes n} : \mathfrak{T}(\mathcal{H}_A^{\otimes n}) \rightarrow \mathfrak{T}(\mathcal{H}_B^{\otimes n})$.

Definition 10.19. A number $R \geq 0$ is called the *achievable rate* (for the transmission of quantum information), if there exist a sequence of spaces $\mathcal{H}^{(n)}$, such that

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \dim \mathcal{H}^{(n)} = R,$$

and sequences of channels $\mathcal{E}^{(n)} : \mathfrak{T}(\mathcal{H}^{(n)}) \rightarrow \mathfrak{T}(\mathcal{H}_A^{\otimes n})$ (encodings) and $\mathcal{D}^{(n)} : \mathfrak{T}(\mathcal{H}_B^{\otimes n}) \rightarrow \mathfrak{T}(\mathcal{H}^{(n)})$ (decodings) such that

$$\lim_{n \rightarrow \infty} F_s(\mathcal{H}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) = 1. \quad (10.34)$$

The least upper bound of achievable rates will be denoted $Q(\Phi)$ and called the *quantum capacity* of the channel Φ .

The following data-processing inequality is useful for obtaining bounds on the quantum capacity.

Proposition 10.20. *For any two channels Φ_1, Φ_2*

$$Q(\Phi_2 \circ \Phi_1) \leq \min \{Q(\Phi_1), Q(\Phi_2)\}. \quad (10.35)$$

In particular, if one of these channels has zero quantum capacity, so does their concatenation.

Proof. To show that $Q(\Phi_2 \circ \Phi_1) \leq Q(\Phi_2)$, it suffices to observe that $Q(\Phi_2 \circ \Phi_1)$ is equal to the supremum of the achievable rates for channel Φ_2 over the special class of encodings, including postprocessing with channel Φ_1 . Hence, it does not exceed $Q(\Phi_2)$. The inequality $Q(\Phi_2 \circ \Phi_1) \leq Q(\Phi_1)$ is proved similarly by considering decodings for Φ_1 , including preprocessing with channel Φ_2 . \square

Note that essentially the same reasoning implies a similar inequality for the classical capacity,

$$C(\Phi_2 \circ \Phi_1) \leq \min \{C(\Phi_1), C(\Phi_2)\}, \quad (10.36)$$

while for the classical entanglement-assisted capacity C_{ea} this can be derived from the data-processing inequality for quantum mutual information (Proposition 7.31, property iv), taking into account expression (9.4).

In the proof of the classical Coding Theorem it was much more convenient to use the average error probability instead of the maximal error probability, as the two criteria were shown to be equivalent in Lemma 4.11. In the quantum case, the role of the average error is played by the quantity $1 - F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)})$, where $\bar{S}^{(n)} = \frac{1}{d_n} I_{\mathcal{H}^{(n)}}$ is the chaotic state in $\mathcal{H}^{(n)}$, while the analog of the maximal error is $1 - F_s(\mathcal{H}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)})$. Indeed, $F_s(\mathcal{H}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \rightarrow 1$ implies $F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \rightarrow 1$ (according to Lemma 10.12). In the reverse direction, there is the following noncommutative analog of Lemma 4.11:

Lemma 10.21. *Let \mathcal{H} be a Hilbert space of dimensionality $2d$ and Φ a channel in \mathcal{H} . There exists a subspace \mathcal{H}' of dimensionality d such that*

$$1 - F_s(\mathcal{H}', \mathcal{D}' \circ \Phi) \leq 2(1 - F_e(\bar{S}, \Phi)),$$

where $\bar{S} = \frac{1}{2d} I_{\mathcal{H}}$, and $\mathcal{D}'[S] = P' S P' + \frac{P'}{d} \text{Tr}(I - P') S$ for all $S \in \mathfrak{S}(\mathcal{H})$ with P' being the projector onto \mathcal{H}' .

Proof. On the unit ball of \mathcal{H} , consider the continuous function $|\psi\rangle \rightarrow f(\psi) = \langle \psi | \Phi[|\psi\rangle\langle\psi|] |\psi\rangle = F_e(|\psi\rangle\langle\psi|, \Phi)$. Let $|\psi_1\rangle$ be a vector minimizing $f(\psi)$. Define the orthonormal basis $\{|\psi_j\rangle; j = 1, \dots, 2d\}$ in $\mathcal{H} \equiv \mathcal{H}_0$ by the following recurrent procedure. $|\psi_{j+1}\rangle$ is the vector in the subspace $\mathcal{H}_j = \{|\psi_1\rangle, \dots, |\psi_j\rangle\}^\perp$ that minimizes $f(\psi)$. Then $\mathcal{H}_j \supset \mathcal{H}_{j+1}$ and $\dim \mathcal{H}_j = 2d - j$. By the convexity of $F_e(S, \Phi)$ (see Exercise 10.11), we obtain

$$\begin{aligned} 1 - F_e(\bar{S}, \Phi) &\geq \frac{1}{2d} \sum_{j=1}^{2d} (1 - F_e(|\psi_j\rangle\langle\psi_j|, \Phi)) \\ &\geq \frac{1}{2d} \sum_{j=d+1}^{2d} (1 - f(\psi_j)) \\ &\geq \frac{1}{2} \left(1 - \min_{|\psi\rangle \in \mathcal{H}_d} f(\psi) \right) \\ &\geq \frac{1}{2} (1 - F_s(\mathcal{H}', \mathcal{D}' \circ \Phi)), \end{aligned}$$

where $\mathcal{H}' = \mathcal{H}_d$. □

From this lemma, together with inequality (10.26), it follows that in the definition of achievable rates we can replace (10.34) by

$$\lim_{n \rightarrow \infty} F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) = 1. \quad (10.37)$$

Corollary 10.22. *Any PPT channel (in particular, an entanglement-breaking channel) has zero quantum capacity.*

Proof. In this proof, we denote by T the operations of transposition in different spaces, which will be clear from the context. By using (6.31), we have

$$\begin{aligned} F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) &= F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ T \circ T \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \\ &= F_e(\bar{S}^{(n)}, T \circ \Phi_n), \end{aligned}$$

where $\Phi_n = (\mathcal{D}^{(n)})^\top \circ (T \circ \Phi^{\otimes n}) \circ \mathcal{E}^{(n)}$ is a channel, because by Proposition 6.27 $T \circ \Phi^{\otimes n}$ is a channel. Now,

$$\begin{aligned} F_e(\bar{S}^{(n)}, T \circ \Phi_n) &= \langle \Omega^{(n)} | (\text{Id}_{\mathcal{H}^{(n)}} \otimes T \circ \Phi_n) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right] |\Omega^{(n)}\rangle \\ &= \text{Tr} |\Omega^{(n)}\rangle\langle\Omega^{(n)}| (\text{Id}_{\mathcal{H}^{(n)}} \otimes T \circ \Phi_n) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right] \\ &= \text{Tr} (\text{Id}_{\mathcal{H}^{(n)}} \otimes T) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right] (\text{Id}_{\mathcal{H}^{(n)}} \otimes \Phi_n) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right], \end{aligned}$$

where

$$|\Omega^{(n)}\rangle = \frac{1}{\sqrt{d_n}} \sum_{m=1}^{d_n} |m\rangle \otimes |m\rangle \quad (10.38)$$

is the maximally entangled vector in $\mathcal{H}^{(n)} \otimes \mathcal{H}^{(n)}$.

Introducing the Kraus decomposition $\Phi_n[S] = \sum_k V_k S V_k^*$, we obtain, by using (10.38), that the last expression is evaluated as

$$\begin{aligned} \frac{1}{d_n^2} \sum_k \sum_{l,m=1}^{d_n} \langle m|V_k|l\rangle \langle m|V_k^*|l\rangle &\leq \frac{1}{d_n^2} \sum_k \sum_{l,m=1}^{d_n} |\langle m|V_k|l\rangle|^2 = \frac{1}{d_n^2} \sum_k \text{Tr } V_k^* V_k \\ &= \frac{1}{d_n}. \end{aligned}$$

Thus, for arbitrary encodings and decodings $F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) = d_n^{-1} \leq 1/2$ implying $Q(\Phi) = 0$. \square

Now, let us show that in the definition of the quantum capacity we can restrict ourselves to *isometric* encodings, i.e. to those which have the form

$$\mathcal{E}[S] = \text{Ad}V[S] \equiv V S V^*, \quad (10.39)$$

where V is an isometric mapping of the space $\mathcal{H}^{(n)}$ into the input space $\mathcal{H}^{\otimes n}$ of the channel. We shall prove that if there is an encoding of a general form which achieves high fidelity for a given state, there also is an isometric encoding with a similar property. To explain this on an intuitive level, consider the case of perfect transmission. If the concatenation encoding-channel-decoding perfectly reproduces some state S , the encoding channel \mathcal{E} is perfectly reversible on S , with the concatenation channel-decoding as the reverse channel. In this case, according to Proposition 10.7 the channel \mathcal{E} on $\text{supp } S$ can be represented in the form (10.5) as a convex combination of isometric encodings $\mathcal{U}_j = \text{Ad}U_j$. Hence, any channel \mathcal{U}_j is perfectly reversible on S with the same reverse channel.

Lemma 10.23. *Let S be a state in a Hilbert space \mathcal{L} , $\mathcal{L} = \text{supp } S$, let \mathcal{E} be a completely positive map from $\mathfrak{T}(\mathcal{L})$ to $\mathfrak{T}(\mathcal{H})$, such that $\text{Tr } \mathcal{E}[S] = 1$, and let \mathcal{A} be a channel from $\mathfrak{T}(\mathcal{H})$ to $\mathfrak{T}(\mathcal{L})$. Assume that $\dim \mathcal{L} \leq \dim \mathcal{H}$. Then there exists an isometry V from \mathcal{L} to \mathcal{H} such that*

$$F_e(S, \mathcal{A} \circ \text{Ad}V) \geq F_e(S, \mathcal{A} \circ \mathcal{E})^2. \quad (10.40)$$

This lemma will be used in the situation where \mathcal{E} is the encoding map, while \mathcal{A} is the concatenation channel-decoding. For the proof, we need the following generalization of the polar decomposition. If X is an operator from \mathcal{L} to \mathcal{H} , $\dim \mathcal{L} \leq \dim \mathcal{H}$, then $X = V|X|$, where $|X| = \sqrt{X^*X}$ is a positive operator in \mathcal{L} and V is an isometry from \mathcal{L} to \mathcal{H} . This implies the following representation for an arbitrary square

matrix X : $X = VDU$, where D is a diagonal matrix with nonnegative elements, and V, U are unitary matrices.

Proof. Let $A_i : \mathcal{H} \rightarrow \mathcal{L}$ and $E_j : \mathcal{L} \rightarrow \mathcal{H}$ be the components of the Kraus decomposition for the maps \mathcal{A} and \mathcal{E} , respectively. Denote by X the matrix with the elements $X_{ij} = \text{Tr } A_i E_j S$. Then, by (10.21),

$$F_e(S, \mathcal{A} \circ \mathcal{E}) = \sum_{ij} |X_{ij}|^2.$$

Complementing the Kraus decompositions, if necessary, with additional zero components, we may assume that X is a square matrix. The decomposition $X = VDU$ implies that by transforming the Kraus decompositions for \mathcal{A} and \mathcal{E} , we can make the matrix X diagonal. Then again using (10.21), we obtain $F_e(S, \mathcal{A} \circ \mathcal{E}) = \sum_k |\text{Tr } A_k E_k S|^2$. Denote $\lambda_k = \text{Tr } S E_k^* E_k$, and restrict ourselves to $\lambda_k > 0$. Then $\sum_k \lambda_k (|\text{Tr } A_k E_k S|^2 / \lambda_k) = F_e(S, \mathcal{A} \circ \mathcal{E})$ and $\sum_k \lambda_k = 1$, so that there exists a k such that $|\text{Tr } A_k E_k S|^2 / \lambda_k \geq F_e(S, \mathcal{A} \circ \mathcal{E})$. Setting $E = E_k / \sqrt{\lambda_k} : \mathcal{L} \rightarrow \mathcal{H}$, $A = A_k : \mathcal{H} \rightarrow \mathcal{L}$, we have $A^* A \leq I_{\mathcal{H}}$, $\text{Tr } S E^* E = 1$ and

$$F_e(S, \mathcal{A} \circ \mathcal{E}) \leq F_e(S, \text{Ad } A \circ \text{Ad } E) = |\text{Tr } A E S|^2.$$

Let $A^* = V|A^*|$ be the polar decomposition of the operator $A^* : \mathcal{L} \rightarrow \mathcal{H}$, where $V : \mathcal{L} \rightarrow \mathcal{H}$ is an isometry. According to the noncommutative Cauchy–Schwarz inequality (2.24)

$$|\text{Tr } A E S|^2 = |\text{Tr } S A E|^2 \leq \text{Tr } S A A^* \text{Tr } S E^* E = \text{Tr } S |A^*|^2. \quad (10.41)$$

Since $A^* A \leq I_{\mathcal{H}}$, also $A A^* \leq I_{\mathcal{L}}$. Hence, $|A^*|^2 \leq |A^*|$ and therefore

$$\text{Tr } S |A^*|^2 \leq \text{Tr } S |A^*| = \text{Tr } A V S.$$

Then

$$|\text{Tr } A V S|^2 \leq \sum_k |\text{Tr } A_k V S|^2 = F_e(S, \mathcal{A} \circ \text{Ad } V).$$

This chain of inequalities implies (10.40). \square

Now, let some sequences of subspaces $\mathcal{H}^{(n)}$, encodings $\mathcal{E}^{(n)}$, and decodings $\mathcal{D}^{(n)}$ be given, for which (10.37) holds. In this case, by Lemma 10.23, there exists a sequence of isometric encodings $\mathcal{V}^{(n)} = \text{Ad } V^{(n)}$ such that

$$\lim_{n \rightarrow \infty} F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{V}^{(n)}) = 1. \quad (10.42)$$

It follows that in the definition of the quantum capacity we can restrict ourselves to the isometric encodings.

10.3.2 The quantum capacity and the coherent information

The following fundamental result is the Coding Theorem for the quantum capacity, which relates it to the coherent information.

Theorem 10.24 (Lloyd; Shor; Devetak). *For any channel Φ , the quantum capacity is given by the expression*

$$Q(\Phi) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_S I_c(S, \Phi^{\otimes n}), \quad (10.43)$$

where $I_c(S, \Phi)$ is the coherent information.

Existence of the limit can be shown, just like in the case of the classical capacity, by using superadditivity of the sequence $\bar{Q}_n = \max_S I_c(S, \Phi^{\otimes n})$. Let us check that (10.43) implies the inequality relating the quantum and the classical capacities of the channel

$$Q(\Phi) \leq C(\Phi), \quad (10.44)$$

which is proved the same way as the second inequality in (9.8). Indeed, take S to be a mixture of arbitrary pure states S_j with probabilities p_j . In this case, (7.50) implies

$$I_c(S, \Phi) \leq H\left(\sum_j p_j \Phi[S_j]\right) - \sum_j p_j H(\Phi[S_j]) = \chi(\{p_j\}, \{\Phi[S_j]\}).$$

Taking the maximum provides $\max_S I_c(S, \Phi) \leq C_\chi(\Phi)$. Applying this to $\Phi^{\otimes n}$ leads to (10.44).

Proof of the inequality \leq . Denote by $\bar{Q}(\Phi)$ the right-hand side of (10.43). A simpler part of the theorem is the proof of the inequality

$$Q(\Phi) \leq \bar{Q}(\Phi), \quad (10.45)$$

i.e. the weak converse of the Coding Theorem. Let $\mathcal{H}^{(n)}$ be the input space of dimensionality $d_n = \dim \mathcal{H}^{(n)} = 2^{nR}$ and let $\mathcal{E}^{(n)}, \mathcal{D}^{(n)}$ be an encoding and decoding such that

$$1 - F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \leq \varepsilon,$$

where $\bar{S}^{(n)} = \frac{1}{d_n} I_{\mathcal{H}^{(n)}}$ is the chaotic state in $\mathcal{H}^{(n)}$, with the entropy $H(\bar{S}^{(n)}) = \log d_n = nR$. Here, we are using the requirement (10.37) in the definition of the achievable rate. According to Lemma 10.23, we can assume that $\mathcal{E}^{(n)}$ are *isometric* encodings. Since

$$\begin{aligned} F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \\ = \langle \Omega^{(n)} | \left(\text{Id}_{\mathcal{H}^{(n)}} \otimes \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)} \right) \left[|\Omega^{(n)}\rangle \langle \Omega^{(n)}| \right] | \Omega^{(n)} \rangle, \end{aligned}$$

where $|\Omega^{(n)}\rangle$ is the maximally entangled vector, it follows by Lemma 10.9 that for the maximally entangled state $|\Omega^{(n)}\rangle\langle\Omega^{(n)}|$ in $\mathcal{H}^{(n)} \otimes \mathcal{H}^{(n)}$ the following holds,

$$\left\| |\Omega^{(n)}\rangle\langle\Omega^{(n)}| - (\text{Id}_{\mathcal{H}^{(n)}} \otimes \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right] \right\|_1 \leq 2\sqrt{\varepsilon}. \quad (10.46)$$

This means that the rate R is achievable for asymptotically perfect transmission of the chaotic state in $\mathcal{H}^{(n)}$.

Denoting $S^{(n)} = \mathcal{E}^{(n)} \left[\bar{S}^{(n)} \right]$ and using the chain rule (10.6) for the coherent information, we have

$$I_c(S^{(n)}, \Phi^{\otimes n}) \geq I_c(S^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n}) = H(S_{B'}) - H(S_{B'R'}), \quad (10.47)$$

where $S_{B'R'} = \left(\text{Id}_{\mathcal{H}^{(n)}} \otimes (\mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \right) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right]$, so that

$$S_{B'} = (\mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \left[\bar{S}^{(n)} \right].$$

The equality $H(S_{B'R'}) = H(S^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n})$ follows from the fact that, due to isometric nature of the encoding $\mathcal{E}^{(n)}$, the state $\left(\text{Id}_{\mathcal{H}^{(n)}} \otimes \mathcal{E}^{(n)} \right) \left[|\Omega^{(n)}\rangle\langle\Omega^{(n)}| \right]$ is pure and can be regarded as a purification of $S^{(n)}$. According to (10.46)

$$\left\| \bar{S}^{(n)} - S_{B'} \right\|_1 \leq \left\| |\Omega^{(n)}\rangle\langle\Omega^{(n)}| - S_{B'R'} \right\|_1 \leq 2\sqrt{\varepsilon}.$$

Here the first inequality follows from a simple observation, to be proved in the following exercise.

Exercise 10.25. Prove the following statement: the trace norm does not increase under partial trace, i.e.

$$\left\| \text{Tr}_{\mathcal{H}_0} T \right\|_1 \leq \|T\|_1,$$

for any positive operator $T \in \mathcal{H} \otimes \mathcal{H}_0$. *Hint:* Use expression (1.19) for the trace norm.

By twice using the estimate (7.26) for the continuity of the entropy, we have

$$\begin{aligned} H(S_{B'}) - H(S_{B'R'}) &= H(\bar{S}^{(n)}) + [H(S_{B'}) - H(\bar{S}^{(n)})] \\ &\quad + [H(|\Omega^{(n)}\rangle\langle\Omega^{(n)}|) - H(S_{B'R'})] \\ &\geq H(\bar{S}^{(n)}) - 6 \log \dim \mathcal{H}^{(n)} \sqrt{\varepsilon} - \frac{2 \log e}{e} \\ &= nR (1 - 6\sqrt{\varepsilon}) - \frac{2 \log e}{e}. \end{aligned}$$

Therefore, taking into account (10.47),

$$\bar{Q}_n = \max_{S^{(n)}} I_c(S^{(n)}, \Phi^{\otimes n}) \geq nR(1 - 6\sqrt{\varepsilon}) - \frac{2\log e}{e},$$

whence $R \leq \lim_{n \rightarrow \infty} \frac{1}{n} \bar{Q}_n = \bar{Q}(\Phi)$.

This completes the proof of inequality (10.45), i.e. the Weak Converse for the quantum capacity. The proof of the Direct Coding Theorem is postponed until Section 10.4.4. \square

10.3.3 Degradable channels

First of all, we obtain important relations between the coherent information and the quantity

$$\chi(\{\pi_x\}, \{S_x\}) = H\left(\sum_x \pi_x S_x\right) - \sum_x \pi_x H(S_x),$$

giving the upper bound (5.16) for the amount of classical information and, as a consequence, a relation between the quantum and the classical capacities of a channel and its complementary.

Consider a channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$, its Stinespring dilation with the isometry $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$, and the complementary channel (see Section 6.6)

$$\tilde{\Phi}[S] = \text{Tr}_{\mathcal{H}_B} V S V^*; \quad S \in \mathfrak{T}(\mathcal{H}_A). \quad (10.48)$$

Let $S = \sum_x \pi_x S_x$ be an arbitrary decomposition of the state S into pure states. In this case, the following equalities hold,

$$\begin{aligned} I_c(S, \Phi) &= H(\Phi[S]) - H(\tilde{\Phi}[S]) \\ &= \left[H(\Phi[S]) - \sum_x \pi_x H(\Phi[S_x]) \right] \end{aligned} \quad (10.49)$$

$$\begin{aligned} &- \left[H(\tilde{\Phi}[S]) - \sum_x \pi_x H(\tilde{\Phi}[S_x]) \right] \\ &= \chi(\{\pi_x\}, \{\Phi[S_x]\}) - \chi(\{\pi_x\}, \{\tilde{\Phi}[S_x]\}) \end{aligned} \quad (10.50)$$

$$= \sum_x \pi_x [H(\Phi[S_x]; \Phi[S]) - H(\tilde{\Phi}[S_x]; \tilde{\Phi}[S])]. \quad (10.51)$$

Here, equality (10.49) follows from the fact that the state $V S_x V^*$ is pure and hence

$$H(\Phi[S_x]) = H(\tilde{\Phi}[S_x])$$

for all x , and (10.51) follows from identity (7.19). Taking the upper and the lower bounds in (10.50), with respect to all possible ensembles of pure states, we obtain

$$-Q_1(\tilde{\Phi}) \leq C_\chi(\Phi) - C_\chi(\tilde{\Phi}) \leq Q_1(\Phi), \quad (10.52)$$

where we introduced the notation

$$Q_1(\Phi) = \max_S I_c(S, \Phi). \quad (10.53)$$

Applying (10.52) to channel $\Phi^{\otimes n}$, taking the limit $n \rightarrow \infty$, and using the Coding Theorems, we obtain,

$$-Q(\tilde{\Phi}) \leq C(\Phi) - C(\tilde{\Phi}) \leq Q(\Phi).$$

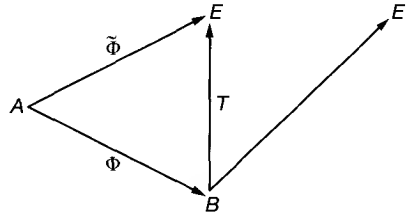


Figure 10.1. Degradable channel.

Now, we introduce an important class of channels for which the quantum capacity is given by the “one-letter” expression (10.53).

Definition 10.26. The channel Φ is called *degradable* if there exists a channel Γ such that $\tilde{\Phi} = \Gamma \circ \Phi$, and *anti-degradable* if there exists a channel Γ' such that $\Phi = \Gamma' \circ \tilde{\Phi}$.

Obviously, Φ is degradable if and only if $\tilde{\Phi}$ is anti-degradable. From relation (10.51), by using the monotonicity of the relative entropy, we deduce that *if Φ is an anti-degradable channel, then*

$$I_c(S, \Phi) \leq 0 \quad \text{for all states } S \quad (10.54)$$

(correspondingly, $I_c(S, \Phi) \geq 0$ for a degradable channel).

Proposition 10.27. *If Φ is an anti-degradable channel, then $Q(\Phi) = 0$. If Φ is degradable, then*

$$Q(\Phi) = Q_1(\Phi) = \max_S I_c(S, \Phi). \quad (10.55)$$

Proof. The first statement follows from (10.54) and the Weak Converse (10.45).

Let Φ be a degradable channel and $W : \mathcal{H}_B \rightarrow \mathcal{H}_E \otimes \mathcal{H}_{E'}$ be the Stinespring isometry for the channel $\Gamma : \mathfrak{T}(\mathcal{H}_B) \rightarrow \mathfrak{T}(\mathcal{H}_E)$. Then $I_c(S, \Phi) = H(B) - H(E) = H(E E') - H(E)$, so that

$$I_c(S, \Phi) = H(E'|E).$$

For two degradable channels Φ_1, Φ_2 and for an arbitrary state S_{12} in $\mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$, we have

$$I_c(S_{12}, \Phi_1 \otimes \Phi_2) = H(E'_1 E'_2 | E_1 E_2),$$

Hence, by subadditivity (7.43) of the conditional entropy

$$I_c(S_{12}, \Phi_1 \otimes \Phi_2) \leq I_c(S_1, \Phi_1) + I_c(S_2, \Phi_2),$$

implying

$$I_c(S^{(n)}, \Phi^{\otimes n}) = n \max_S I_c(S, \Phi) \quad (10.56)$$

and hence (10.55). \square

Consider an entanglement-breaking channel

$$\Phi[S] = \sum_{k=1}^{\tilde{d}} |\varphi_k\rangle\langle\psi_k| S |\psi_k\rangle\langle\varphi_k|,$$

where $\{\psi_k\}$ is an overcomplete system in \mathcal{H} , and $\{\varphi_k\}$ is a system of unit vectors in the output space \mathcal{H}' . The complementary channel is

$$\tilde{\Phi}[S] = \sum_{k,l=1}^{\tilde{d}} c_{kl} |e_k\rangle\langle\psi_k| S |\psi_l\rangle\langle e_l|,$$

where $c_{kl} = \langle\varphi_l|\varphi_k\rangle$, while $\{e_k\}$ is the standard basis in $\mathbb{C}^{\tilde{d}}$. We have $\Phi = \Gamma' \circ \tilde{\Phi}$, where

$$\Gamma'[S] = \sum_{k=1}^{\tilde{d}} |\varphi_k\rangle\langle e_k| S |e_k\rangle\langle\varphi_k|.$$

This proves

Corollary 10.28. *Any entanglement-breaking channel Φ is anti-degradable.*

The first statement of Proposition 10.27 again implies $Q(\Phi) = 0$, thus reconfirming Corollary 10.22.

Example 10.29. Consider the quantum erasure channels Φ_p . For $q \geq p$,

$$\Phi_q = \Gamma_{(1-q)/(1-p)} \circ \Phi_p,$$

where $\Gamma_\alpha : \mathfrak{T}(\mathcal{H} \oplus \mathbb{C}) \rightarrow \mathfrak{T}(\mathcal{H} \oplus \mathbb{C})$ is the channel acting as

$$\Gamma_\alpha \left[\begin{bmatrix} S & 0 \\ 0 & s \end{bmatrix} \right] = \alpha \begin{bmatrix} S & 0 \\ 0 & s \end{bmatrix} + (1-\alpha) \begin{bmatrix} 0 & 0 \\ 0 & \text{Tr } S + s \end{bmatrix}.$$

Combining this with the fact that $\tilde{\Phi}_p = \Phi_{1-p}$ (see Exercise 6.35), we find that Φ_p is degradable for $p \in [0, 1/2]$ and anti-degradable for $p \in [1/2, 1]$. Therefore, by Theorem 10.27, we obtain

$$Q(\Phi_p) = \begin{cases} (1-2p) \log d, & p \in [0, 1/2]; \\ 0, & p \in [1/2, 1]. \end{cases} \quad (10.57)$$

Here, the first line follows from (10.55) together with the fact that $H(\Phi_p[S]) = (1-p)H(S) + h_2(p)$. Hence, $I_c(S, \Phi_p) = H(\Phi_p[S]) - H(\tilde{\Phi}_p[S]) = (1-p)H(S) - pH(S) = (1-2p)H(S)$. Maximizing over S gives (10.57).

10.4 The private classical capacity and the quantum capacity

10.4.1 The quantum wiretap channel

Consider the situation where the quantum communication channel is accessed by three parties, a sender A , a receiver B and an eavesdropper E . The mathematical model of the *quantum wiretap channel* consists of three Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_E$ and an isometric map $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$, so that the input state S_A is mapped to the state $S_{BE} = \Phi_{BE}[S_A] \equiv VS_AV^*$ of the system BE , with the partial states

$$S_B = \Phi_B[S_A] = \text{Tr}_E VS_AV^*, \quad S_E = \Phi_E[S_A] = \text{Tr}_B VS_AV^*.$$

Recall that the notation E was used previously for the environment. This is a consistent change, if we regard the whole environment as accessible to the eavesdropper.

Assume that the party A chooses an ensemble of states $\{S_A^x\}$ with probabilities $\{\pi_x\}$, so that B and E receive states $\{S_B^x\}$ and $\{S_E^x\}$, respectively. In this case, the upper bounds for the Shannon informations of B and E are, correspondingly, $\chi(\{\pi_x\}, \{S_B^x\})$ and $\chi(\{\pi_x\}, \{S_E^x\})$. By analogy with the classical wiretap channel, see Section 4.5, the “quantum privacy” can be characterized by the quantity $\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\})$. Assuming that the input states S_A^x are *pure* and denoting by $\bar{S}_A = \sum_x \pi_x S_A^x$ the average of the input ensemble, we obtain, from (10.50), a key relation (Schumacher and Westmoreland [178]),

$$I_c(\bar{S}_A, \Phi_B) = \chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}). \quad (10.58)$$

This points to the profound connection between the quantum capacity and the private classical capacity, which provides a hint to a proof of the Direct Coding Theorem.

To define the private classical capacity, let us consider block coding for the quantum wiretap channel. The goal is to transmit the maximum amount of classical information from A to B in a private way, so that E can receive only an asymptotically negligible amount of information. A *code* $(\Sigma^{(n)}, M^{(n)})$ of length n and of size N is defined as

in Definition 8.1, i.e. as a collection of states $\Sigma^{(n)} = \{S_{A^{(n)}}^i; i = 1, \dots, N\}$ in $\mathcal{H}_A^{\otimes n}$, together with an observable $M^{(n)} = \{M_j; j = 0, 1, \dots, N\}$ in $\mathcal{H}_B^{\otimes n}$. Along with its error probability $P_e(\Sigma^{(n)}, M^{(n)})$, defined by (8.1), we consider the new quantity

$$v_E(\Sigma^{(n)}) = \max_{i,k=1,\dots,N} \|S_{E^{(n)}}^i - S_{E^{(n)}}^k\|_1,$$

which characterizes the *variability*, and hence the information content of the output states for the eavesdropper E . Notice that, by the triangle inequality for the norm, we have

$$\|\bar{S}_{E^{(n)}} - S_{E^{(n)}}^i\|_1 \leq v_E(\Sigma^{(n)}) \quad \text{for all } i,$$

where $\bar{S}_{E^{(n)}} = \frac{1}{N} \sum_{i=1}^N S_{E^{(n)}}^i$. Applying the inequality (7.25), we obtain

$$\begin{aligned} \frac{1}{n} \chi\left(\left\{\frac{1}{N}\right\}, \{S_{E^{(n)}}^i\}\right) &= \frac{1}{nN} \sum_{i=1}^N \left[H(\bar{S}_{E^{(n)}}) - H(S_{E^{(n)}}^i) \right] \\ &\leq \log \dim \mathcal{H}_E \cdot v_E(\Sigma^{(n)}) + \frac{1}{n} \eta(v_E(\Sigma^{(n)})), \end{aligned} \quad (10.59)$$

provided that $v_E(\Sigma^{(n)})$ is small enough. Therefore, if the variability is small, the Shannon mutual information between A and E , with equiprobable encoding, is also small.

We call R the *achievable rate* for the wiretap channel if there exists a sequence of codes $(\Sigma^{(n)}, M^{(n)})$ of sizes $N = 2^{nR}$, such that

$$\lim_{n \rightarrow \infty} P_e(\Sigma^{(n)}, M^{(n)}) = 0$$

and

$$\lim_{n \rightarrow \infty} v_E(\Sigma^{(n)}) = 0. \quad (10.60)$$

The least upper bound of the achievable rates is called the *private classical capacity* $C_p(\Phi_{BE})$ of the wiretap channel. By (10.59), condition (10.60) implies asymptotic vanishing of the mutual information between A and E . It is possible to show that these conditions are, in fact, equivalent, but condition (10.60) is more convenient technically and also is more useful for later application to the quantum capacity.

Theorem 10.30 (Devetak [51]; Cai, Winter and Yeung [34]).

$$C_p(\Phi_{BE}) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{\pi^{(n)}, \Sigma^{(n)}} \left[\chi(\{\pi_i^{(n)}\}, \{S_{B^{(n)}}^i\}) - \chi(\{\pi_i^{(n)}\}, \{S_{E^{(n)}}^i\}) \right], \quad (10.61)$$

where the maximization is over all ensembles, i.e. finite collections of states $\Sigma^{(n)} = \{S_{A^{(n)}}^i\}$ in $\mathcal{H}_A^{\otimes n}$ with probability distributions $\pi^{(n)} = \{\pi_i^{(n)}\}$ (we use the notations $S_{B^{(n)}}^i = \Phi_B^{\otimes n}[S_{A^{(n)}}^i]$, $S_{E^{(n)}}^i = \Phi_E^{\otimes n}[S_{A^{(n)}}^i]$).

The formulas (10.43), (10.58), and (10.61) imply an important relation between the quantum and the private classical capacities

$$Q(\Phi_B) \leq C_p(\Phi_{BE}) \quad (10.62)$$

following from the fact that $C_p(\Phi_{BE})$ involves arbitrary state ensembles for A , while $Q(\Phi_B)$ involves only pure ones. In general, the inequality can be strict. Therefore, the following statement is of special interest.

Proposition 10.31. *If the channel Φ_B is degradable,*

$$C_p(\Phi_{BE}) = Q(\Phi_B) = Q_1(\Phi_B). \quad (10.63)$$

Proof. By using equality (10.50), we have

$$\begin{aligned} \chi(\{\pi_i^{(n)}\}, \{S_{B^{(n)}}^i\}) - \chi(\{\pi_i^{(n)}\}, \{S_{E^{(n)}}^i\}) \\ &= H\left(\sum_i \pi_i^{(n)} S_{B^{(n)}}^i\right) - H\left(\sum_i \pi_i^{(n)} S_{E^{(n)}}^i\right) \\ &\quad - \sum_i \pi_i^{(n)} \left[H(S_{B^{(n)}}^i) - H(S_{E^{(n)}}^i) \right] \\ &= I_c\left(\sum_i \pi_i^{(n)} S_{A^{(n)}}^i, \Phi_B^{\otimes n}\right) - \sum_i \pi_i^{(n)} I_c(S_{A^{(n)}}^i, \Phi_B^{\otimes n}) \\ &\leq I_c\left(\sum_i \pi_i^{(n)} S_{A^{(n)}}^i, \Phi_B^{\otimes n}\right), \end{aligned}$$

since channel $\Phi_B^{\otimes n}$, along with Φ_B , is degradable and hence, $I_c(S_{A^{(n)}}^i, \Phi_B^{\otimes n}) \geq 0$. Therefore, by using (10.62)

$$\max_{\pi^{(n)}, \Sigma^{(n)}} \left[\chi(\{\pi_i^{(n)}\}, \{S_B^i\}) - \chi(\{\pi_i^{(n)}\}, \{S_{E^{(n)}}^i\}) \right] = \max_{S^{(n)}} I_c(S^{(n)}, \Phi_B^{\otimes n}),$$

which is equal to $nQ_1(\Phi_B)$ due to (10.56). Substituting this into (10.61) we obtain (10.63). \square

10.4.2 Proof of the Private Capacity Theorem

We first prove inequality \leq i.e. the Weak Converse. Let R be an achievable rate. In this case, by (5.33) and the information bound (5.16),

$$\begin{aligned} \bar{P}_e(\Sigma^{(n)}, M^{(n)}) &\geq 1 - \frac{\chi(\{1/N\}, \{S_{B^{(n)}}^i\})}{nR} - \frac{1}{nR} \\ &= 1 - \frac{\chi(\{1/N\}, \{S_{B^{(n)}}^i\}) - \chi(\{1/N\}, \{S_{E^{(n)}}^i\})}{nR} \\ &\quad - \frac{1 + \chi(\{1/N\}, \{S_{E^{(n)}}^i\})}{nR}. \end{aligned} \quad (10.64)$$

According to (10.59),

$$\chi(\{1/N\}, \{S_{E^{(n)}}^i\}) \leq v_E(\Sigma^{(n)}) \left[n \log d_E - \log v_E(\Sigma^{(n)}) \right],$$

where $d_E = \dim \mathcal{H}_E$. Therefore, the last term in the right hand side of (10.64) tends to zero as $n \rightarrow \infty$, along with the left hand side. From this follows $0 \geq 1 - \bar{C}_p/R$, where \bar{C}_p is the right hand side of (10.61), and hence $C_p(\Phi_{BE}) \leq \bar{C}_p$.

To prove the direct statement it is sufficient to show that for an arbitrary ensemble $\{\pi_x, S_A^x\}$ and $\delta > 0$ the rate $\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}) - \delta$ is achievable. Achievability of the rate $\bar{C}_p - \delta$ then follows by additional blocking. Consider the c-q channel $x \rightarrow S_B^x$ and the random codebook $W^{(n)}$ of size $N = 2^{n[\chi(\{\pi_x\}, \{S_B^x\}) - c\delta]}$, with independent words having the probability distribution

$$P\{w = (x_1, \dots, x_n)\} = \pi_{x_1} \cdots \pi_{x_n}. \quad (10.65)$$

The constant c in the definition of N will be fixed later. To simplify the notation, we will denote all constants that appear below by the same letter c . In fact, we can always take the maximum of them. By the remark that follows the proof of the Direct Quantum Coding Theorem at the end of Section 5.6, for sufficiently large n , there exist decision rules $M^{(n)}$ for the system B such that

$$\mathbb{E} \bar{P}_e(W^{(n)}, M^{(n)}) \leq 2^{-n\beta}, \quad \beta > 0.$$

Now, consider the new random code with independent words having the uniform distribution on the set $\hat{T}^{n,\delta}$ of δ -strongly typical words,

$$\tilde{P}(w) = \begin{cases} \frac{1}{|\hat{T}^{n,\delta}|}, & w \in \hat{T}^{n,\delta}; \\ 0, & w \notin \hat{T}^{n,\delta}. \end{cases}$$

By Exercise 9.10, $|\hat{T}^{n,\delta}| \geq 2^{n[H(\pi) - \Delta_n(\delta)]}$, where $\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \Delta_n(\delta) = 0$. Since $P(w) \geq 2^{-n[H(\pi) + c\delta]}$ for $w \in \hat{T}^{n,\delta}$, where c is a constant (depending on $\{\pi_x\}$), we have $\tilde{P}(w) \leq 2^{n[c\delta + \Delta_n(\delta)]} P(w)$ for all w , and hence

$$\tilde{E}\tilde{P}_e(W^{(n)}, M^{(n)}) \leq 4^{n[c\delta + \Delta_n(\delta)]} 2^{-n\beta} \leq \varepsilon$$

for n large enough, because the estimate (5.52) for $\tilde{P}_e(W^{(n)}, M^{(n)})$ involves at most two independent words. Hence, by the Markov inequality,

$$\tilde{P}\left\{\tilde{P}_e(W^{(n)}, M^{(n)}) \geq \sqrt{\varepsilon}\right\} \leq \sqrt{\varepsilon} \quad (10.66)$$

So far, this was just a slight modification of the random code for the classical information transmission from A to B . To make it secure against eavesdropping, A has to sacrifice $n[\chi(\{\pi_x\}, \{S_E^x\}) + c\delta/2]$ bits of information, by additional randomization of the input. Assuming $\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}) > 0$, set

$$N_E = 2^{n[\chi(\{\pi_x\}, \{S_E^x\}) + c\delta/2]}, \quad N_B = 2^{n[\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}) - 3c\delta/2]},$$

so that $N_E N_B = N$, and arrange the codebook $W^{(n)}$ in a rectangular array with N_B rows and N_E columns. In this case,

$$\begin{aligned} W^{(n)} &= \{w^{mj}; m = 1, \dots, N_B; j = 1, \dots, N_E\}, \\ M^{(n)} &= \{M^{mj}; m = 1, \dots, N_B; j = 1, \dots, N_E\}. \end{aligned}$$

(For brevity of notation we omit the zero component of $M^{(n)}$ which corresponds to “no decision”).

For each m , A will take the value j at random, with equal probabilities, which results in the input states

$$m \rightarrow S_{A^{(n)}}^m = \frac{1}{N_E} \sum_{j=1}^{N_E} S_{A^{(n)}}^{w^{mj}},$$

with the corresponding outputs $S_{B^{(n)}}^m$ for B and $S_{E^{(n)}}^m$ for E . (Let us recall here, that $S_{A^{(n)}}^w = S_A^{x_1} \otimes \dots \otimes S_A^{x_n}$ for a word $w = (x_1, \dots, x_n)$.) Such a randomization will make the transmitted information almost secret to E , because from the viewpoint of the eavesdropper, for every m the codebook

$$\{w^{mj}; j = 1, \dots, N_E\}$$

with a high probability is a “good” codebook, that carries almost the maximum possible information from A to E , provided an optimal strategy is applied by E . Therefore, the mutual information between the codebooks with different m must be close to zero.

Hence, randomizing inside each of these codebooks discards almost all information carried from A to E .

The crucial step in the rigorous justification of this scenario is the application of the following estimate, based on a quantum version of the Bernstein inequality to be established in the next section.

Proposition 10.32. *Let $S_{E(n)}^{w^{mj}}$; $m = 1, \dots, N_B$; $j = 1, \dots, N_E$ be the random i.i.d. density operators, with N_B, N_E chosen as above. There exists a $\beta_1 > 0$ such that, for n large enough,*

$$\tilde{P} \left\{ \left\| \frac{1}{N_E} \sum_{j=1}^{N_E} S_{E(n)}^{w^{mj}} - \theta \right\|_1 \geq \varepsilon \right\} \leq 2^{-2^{n\beta_1}}, \quad (10.67)$$

for each $m = 1, \dots, N_B$, where θ is a nonrandom operator (in fact, one can take $\theta = \tilde{E} S_{E(n)}^{w^{mj}}$, but we do not need this).

This estimate implies that the probability

$$\begin{aligned} \tilde{P} \left\{ \left\| S_{E(n)}^m - \theta \right\|_1 < \varepsilon; \quad m = 1, \dots, N_B \right\} &\geq 1 - N_B 2^{-2^{n\beta_1}} \\ &= 1 - 2^n [\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}) - 3c\delta/2] - 2^{n\beta_1}, \end{aligned}$$

can be made arbitrarily close to 1 for n large enough. Together with (10.66) this implies that there exists a realization of the codebook $W^{(n)} = \{w^{mj}\}$, for which

$$\bar{P}_e(W^{(n)}, M^{(n)}) < \sqrt{\varepsilon};$$

and

$$\left\| S_{E(n)}^m - S_{E(n)}^l \right\|_1 < 2\varepsilon; \quad m, l = 1, \dots, N_B. \quad (10.68)$$

Defining $\Sigma^{(n)} = \{S_{A(n)}^m; m = 1, \dots, N_B\}$, $\tilde{M}^{(n)} = \{\tilde{M}_m; m = 0, 1, \dots, N_B\}$, where $\tilde{M}_m = \sum_{j=1}^{N_E} M_{mj}$, we have $\bar{P}_e(\Sigma^{(n)}, \tilde{M}^{(n)}) \leq \bar{P}_e(W^{(n)}, M^{(n)}) < \sqrt{\varepsilon}$. An argument similar to Lemma 4.11 shows that we can choose a subcode for which the maximal error $P_e(W^{(n)}, M^{(n)}) < 2\sqrt{\varepsilon}$ while $v_E(\Sigma^{(n)}) < \varepsilon'$, where $\varepsilon' \rightarrow 0$ if $\varepsilon \rightarrow 0$, which proves the theorem. \square

Proof of Proposition 10.32. As in Section 5.6, we introduce the density operator

$$\bar{S}_\pi = \sum_x \pi_x S_E^x,$$

and we write for brevity,

$$H(\bar{S}_\pi) = H(E), \quad \sum_x \pi_x H(S_E^x) = H(E|A),$$

with $\chi(\{\pi_x\}, \{S_E^x\}) = H(E) - H(E|A)$.

Denote by λ_j the eigenvalues of \bar{S}_π . Let us fix a small positive δ , and let $P = P^{n,\delta}$ be the strongly typical projector of the density operator $\bar{S}_\pi^{\otimes n}$, corresponding to the eigenvalues $\lambda_J = \lambda_{j_1} \cdots \lambda_{j_n}$ for which the sequence $J = (j_1, \dots, j_n)$ is strongly typical, i.e. the numbers $N(j|J)$ of times the symbol j appears in J satisfy the condition

$$\left| \frac{N(j|J)}{n} - \lambda_j \right| < \delta, \quad j = 1, \dots, d_E,$$

and $N(j|J) = 0$ if $\lambda_j = 0$ (see Definition 9.9). Denote by $\lambda_j(x)$ the eigenvalues of the density operator S_E^x . For a codeword $w = (x_1, \dots, x_n)$ of length n , let $S_w = S_E^{x_1} \otimes \cdots \otimes S_E^{x_n}$ and denote by P_w the typical projector of S_w , corresponding to the eigenvalues $\lambda_J(w) = \lambda_{j_1}(x_1) \cdots \lambda_{j_n}(x_n)$, for which

$$\left| \frac{N(j, x|J, w)}{n} - \frac{N(x|w)}{n} \lambda_j(x) \right| < \delta, \quad j = 1, \dots, d_E,$$

and $N(j, x|J, w) = 0$ if $\lambda_j(x) = 0$. We shall need the following properties, which are just a reformulation of the corresponding properties for strongly typical sequences in the classical information theory, see [44]:

- i. For $w \in \hat{T}^{n,\delta}$, the eigenvalues of the operator $P_w S_w P_w$ lie in the interval $(2^{-n[H(E|A)+c\delta]}, 2^{-n[H(E|A)-c\delta]})$, where c is another constant, depending on $\{\lambda_j(x)\}$ (to simplify the notation we use the same c as before. In fact, we can choose the maximum of the two constants). In particular, $P_w S_w P_w \leq 2^{-n[H(E|A)-c\delta]} P_w$.
- ii. For $\varepsilon_1 > 0$ and sufficiently large n , $\text{Tr} P_w S_w \geq 1 - \varepsilon_1$. (This is a corollary of the Law of Large Numbers for the probability distribution given by the eigenvalues of S_w).

The following property is less straightforward.

- iii. For given $\varepsilon_1, \delta > 0$, there is a $\delta_1 > 0$ such that for $w \in \hat{T}^{n,\delta_1}$ and sufficiently large n , $\text{Tr} P S_w \geq 1 - \varepsilon_1$, where $P = P^{n,\delta}$ is the strongly typical projector of $\bar{S}_\pi^{\otimes n}$.

Proof. Denote by $p^{(k)}(j) = \langle e_j | S_E^{x_k} | e_j \rangle$, where $|e_j\rangle$ are the eigenvectors of \bar{S}_π , and notice that $p^{(k)}(j) = 0$ if $\lambda_j = 0$, because $\text{supp } S_E^{x_k} \subset \text{supp } \bar{S}_\pi$. In this case,

$$\begin{aligned} \text{Tr } P^{n,\delta} S_w &= \sum_{J \in B^{n,\delta}} p^{(1)}(j_1) \cdots p^{(n)}(j_n) \\ &= \mathbb{P} \left\{ \left| \frac{N(j|J)}{n} - \lambda_j \right| < \delta, \text{ if } \lambda_j > 0 \right\}, \end{aligned} \quad (10.69)$$

where \mathbb{P} is the probability distribution, ascribing the probability $p^{(1)}(j_1) \cdots p^{(n)}(j_n)$ to the sequence $J = (j_1, \dots, j_n)$, and $B^{n,\delta}$ is the set of strongly typical sequences. The fact that $w \in \hat{T}^{n,\delta_1}$ is expressed by similar inequalities

$$\left| \frac{N(x|w)}{n} - \pi_x \right| < \delta_1, \quad \text{if } \pi_x > 0,$$

and $N(x|w) = 0$ if $\pi_x = 0$, which implies for $\bar{\lambda}_j = \frac{1}{n} \sum_{k=1}^n p^{(k)}(j)$,

$$\begin{aligned} |\bar{\lambda}_j - \lambda_j| &= \left| \langle e_j | \left[\frac{1}{n} \sum_{k=1}^n S_E^{x_k} - \bar{S}_\pi \right] | e_j \rangle \right| \\ &= \left| \sum_x \left[\frac{N(x|w)}{n} - \pi_x \right] \langle e_j | S_E^x | e_j \rangle \right| < \delta_1 K, \end{aligned}$$

where K is the number of symbols x . Therefore, by choosing $\delta_1 = \delta/2K$, we have

$$\begin{aligned} \mathbb{P} \left\{ \left| \frac{N(j|J)}{n} - \lambda_j \right| < \delta; \text{ for all } j : \lambda_j > 0 \right\} \\ &\geq \mathbb{P} \left\{ \left| \frac{N(j|J)}{n} - \bar{\lambda}_j \right| < \delta/2; \text{ for all } j : \lambda_j > 0 \right\} \\ &\geq 1 - \sum_{j: \lambda_j > 0} \mathbb{P} \left\{ \left| \frac{N(j|J)}{n} - \bar{\lambda}_j \right| \geq \delta/2 \right\}. \end{aligned}$$

But under the probability distribution \mathbb{P} , the random variable $N(j|J)$ has mean value $\sum_{k=1}^n p^{(k)}(j) = n\bar{\lambda}_j$ and variance $\sum_{k=1}^n p^{(k)}(j)(1 - p^{(k)}(j)) \leq n/4$. Applying the Chebyshev inequality and taking into account (10.69), we obtain iii. \square

We proceed with the proof of Proposition 10.32, for which we also need the following lemma.

Lemma 10.33. *For a positive operator X and a projector P*

$$\|X - PXP\|_1 \leq 3\sqrt{\text{Tr } X \text{Tr } (I - P)X}.$$

Proof. By the triangle inequality and the properties of the trace norm

$$\|X - PXP\|_1 \leq \|(I - P)X(I - P)\|_1 + 2\|PX(I - P)\|_1 \leq 3\|X(I - P)\|_1.$$

There is a unitary U such that $\|X(I - P)\|_1 = \text{Tr } X(I - P)U$. By the Cauchy-Schwarz inequality for the trace (take $S = I/d$ in (2.24))

$$\left[\text{Tr } \sqrt{X} \sqrt{X} (I - P)U \right]^2 \leq \text{Tr } X \text{Tr } U^* (I - P)X(I - P)U = \text{Tr } X \text{Tr } (I - P)X,$$

Hence, the result follows. \square

To simplify notation in (10.67), we shall denote $S_{E(n)}^{w^{mj}} = S_{w^j}$, where the code-words $w^j; j = 1, \dots, N_E$, are i.i.d. uniformly on the set T^{n, δ_1} . We shall obtain (10.67) by applying the operator Bernstein inequality (10.76), to be established in the next section, to the random operators

$$X_j = 2^{n[H(E|A) - \delta]} \Pi P P_{w^j} S_{w^j} P_{w^j} P \Pi,$$

where Π is the projector onto the eigenspace of the operator

$$\theta' = \tilde{E} P P_{w^j} S_{w^j} P_{w^j} P$$

corresponding to eigenvalues $\geq 2^{-n[H(E) + 3c\delta/2]}$. By construction, $0 \leq X_j \leq I$,

$$M = \tilde{E} X_j = 2^{n[H(E|A) - c\delta]} \tilde{E} (\Pi P P_{w^j} S_{w^j} P_{w^j} P \Pi) = 2^{n[H(E|A) - c\delta]} \Pi \theta' \Pi$$

and $\mu = 2^{n[H(E|A) - H(E) - 5c\delta/2]} = 2^{-n[\chi(\{\pi_x\}, \{S_E^x\}) + 5c\delta/2]}$, so that $N_E \mu = 2^{-2nc\delta}$ and (10.76) imply

$$\begin{aligned} \tilde{P} \left\{ \omega : \left\| \frac{1}{N_E} \sum_{j=1}^{N_E} X_j(\omega) - M \right\|_1 \geq \varepsilon_2 2^{n[H(E|A) - c\delta]} \right\} &\leq 2d_E^n \exp(-2^{-2nc\delta} \varepsilon_2^2/4) \\ &\leq 2^{-2^{-n\beta_1}} \end{aligned} \quad (10.70)$$

for some $\beta_1 > 0$ and n large enough. Here, we used the fact that $\text{Tr } M \leq 2^{n[H(E|A) - c\delta]}$. It remains for us to derive (10.67) from (10.70).

Denoting $\theta = \Pi \theta' \Pi$, we have

$$\left\| \frac{1}{N_E} \sum_{j=1}^{N_E} S_{w^j} - \theta \right\|_1 \leq \frac{1}{N_E} \sum_{j=1}^{N_E} \|S_{w^j} - P S'_{w^j} P\|_1 \quad (10.71)$$

$$+ \left\| \frac{1}{N_E} \sum_{j=1}^{N_E} P S'_{w^j} P - \frac{1}{N_E} \sum_{j=1}^{N_E} \Pi P S'_{w^j} P \Pi \right\|_1 \quad (10.72)$$

$$+ \left\| \frac{1}{N_E} \sum_{j=1}^{N_E} \Pi P S'_{w^j} P \Pi - \theta \right\|_1. \quad (10.73)$$

The estimate (10.70) implies

$$\tilde{\mathbb{P}} \left\{ \left\| \frac{1}{N_E} \sum_{j=1}^{N_E} \Pi P S'_{w^j} P \Pi - \theta \right\|_1 \leq \varepsilon_2 \right\} > 1 - 2^{-2^{-n\beta_1}}.$$

The inequality

$$\left\| \frac{1}{N_E} \sum_{j=1}^{N_E} \Pi P S'_{w^j} P \Pi - \theta \right\|_1 \leq \varepsilon_2 \quad (10.74)$$

gives the estimate for the term (10.73).

For the term (10.71), we use the triangle inequality

$$\|S_{w^j} - P S'_{w^j} P\|_1 \leq \|S_{w^j} - P_{w^j} S'_{w^j} P_{w^j}\|_1 + \|S'_{w^j} - P S'_{w^j} P\|_1.$$

The first term on the right can be made small for sufficiently large n by property ii. and Lemma 10.33. Further, notice that for $w^j \in \hat{T}^{n, \delta_1}$

$$\mathrm{Tr} P S'_{w^j} = \mathrm{Tr} S'_{w^j} - \mathrm{Tr} (I - P) S'_{w^j} \geq \mathrm{Tr} P_{w^j} S_{w^j} - \mathrm{Tr} (I - P) S_{w^j} \geq 1 - 2\varepsilon_1 \quad (10.75)$$

by ii. and iii. Applying the lemma, we can also make the second term on the right small for sufficiently large n .

The term (10.72) has the form $\|S - \Pi S \Pi\|_1$. Therefore, to show that it is close to zero, by Lemma 10.33 it is sufficient to show that

$$\mathrm{Tr} \Pi S = \mathrm{Tr} \frac{1}{N_E} \sum_{j=1}^{N_E} \Pi P S'_{w^j} P \Pi$$

is close to 1. But, according to (10.74), this is not less than $\mathrm{Tr} \theta - \varepsilon_2 = \mathrm{Tr} \Pi \theta' - \varepsilon_2$. Now, $\mathrm{Tr} \Pi \theta' = \mathrm{Tr} \theta' - \mathrm{Tr} (I - \Pi) \theta' \geq \mathrm{Tr} \theta' - 2^{-nc\delta/2}$, because $\mathrm{Tr} (I - \Pi) \theta'$ is the sum of eigenvalues of θ' that are less than $2^{-n[H(E)+3c\delta/2]}$, while the total number of positive eigenvalues is less than or equal to $\dim \mathrm{supp} \theta' \leq \dim \mathrm{supp} P = 2^{n[H(E)+c\delta]}$. It remains for us to prove that $\mathrm{Tr} \theta'$ is close to 1. However, $\theta' = \tilde{E} P S'_{w^j} P$ and $\mathrm{Tr} \theta' = \tilde{E} \mathrm{Tr} P S'_{w^j} \geq 1 - 2\varepsilon_1$ by (10.75). This implies that, given ε , we can choose $\varepsilon_1, \varepsilon_2$ such that (10.67) holds for n large enough. \square

10.4.3 Large deviations for random operators

Theorem 10.34 (Ahlsvede and Winter [4]). *Let $X_1(\omega), \dots, X_N(\omega)$ be i.i.d. operator-valued random variables in \mathcal{H} , $\dim \mathcal{H} = d$, such that $0 \leq X_i(\omega) \leq I$, $M = \mathbb{E} X_i \geq \mu I$. Then, for $0 \leq \varepsilon \leq 1$,*

$$\mathbb{P} \left\{ \omega : \left\| \frac{1}{N} \sum_{j=1}^N X_j(\omega) - M \right\|_1 \geq \varepsilon \mathrm{Tr} M \right\} \leq 2d \exp \left(-\frac{N\mu\varepsilon^2}{4} \right). \quad (10.76)$$

Proof. The proof consists of several steps. First, we prove the operator Markov inequality. Let X be a Hermitean operator-valued random variable, $t > 0$. In this case,

$$\mathbb{P} \{ \omega : X(\omega) \not\leq 0 \} \leq \text{Tr } \mathbb{E} \exp tX,$$

where $\{ \omega : X(\omega) \not\leq 0 \} = \overline{\{ \omega : X(\omega) \leq 0 \}}$. Indeed,

$$\begin{aligned} \mathbb{P} \{ \omega : X(\omega) \not\leq 0 \} &= \mathbb{P} \{ \omega : \exp tX(\omega) \not\leq I \} \\ &\leq \mathbb{E} 1_{\{\exp tX \not\leq I\}} \text{Tr } \exp tX \\ &\leq \text{Tr } \mathbb{E} \exp tX, \end{aligned}$$

where the first inequality follows from the fact that $0 \leq Y \not\leq I$ implies $1 \leq \text{Tr } Y$.

Second, by letting $X(\omega) = \sum_{i=1}^N X_i(\omega)$, where X_i are i.i.d. Hermitean operator-valued random variables, we obtain

$$\begin{aligned} \mathbb{P} \left\{ \omega : \sum_{i=1}^N X_i(\omega) \not\leq 0 \right\} &\leq \text{Tr } \mathbb{E} \exp \left(t \sum_{i=1}^N X_i \right) \\ &\leq \mathbb{E} \text{Tr } \exp \left(t \sum_{i=1}^{N-1} X_i \right) \exp tX_N \\ &= \text{Tr } \mathbb{E} \exp \left(t \sum_{i=1}^{N-1} X_i \right) \mathbb{E} \exp tX_N, \end{aligned}$$

where the Golden–Thompson inequality ([27], Theorem IX.3.7) was used:

$$\text{Tr } \exp(A + B) \leq \text{Tr } \exp A \exp B.$$

By inequality (1.18) this is less than or equal to

$$\text{Tr } \mathbb{E} \exp \left(t \sum_{i=1}^{N-1} X_i \right) \| \mathbb{E} \exp tX_N \|.$$

Applying the estimate iteratively, we get

$$\mathbb{P} \left\{ \omega : \sum_{i=1}^N X_i(\omega) \not\leq 0 \right\} \leq d \| \mathbb{E} \exp tX_i \|^N. \quad (10.77)$$

Third, let $X_1(\omega), \dots, X_N(\omega)$ be i.i.d. operator-valued random variables, such that $0 \leq X_i(\omega) \leq I$, $\mathbb{E} X_i = \mu I$, and $0 \leq \mu \leq a \leq 1$, then

$$\mathbb{P} \left\{ \omega : \frac{1}{N} \sum_{i=1}^N X_i(\omega) \not\leq aI \right\} \leq d \exp(-Nh(a; \mu)), \quad (10.78)$$

where $h(a; \mu) = a \ln \frac{a}{\mu} + (1-a) \ln \frac{1-a}{1-\mu}$. Indeed, by using (10.77) for $X_i(\omega) - aI$, we obtain

$$P \left\{ \omega : \frac{1}{N} \sum_{i=1}^N X_i(\omega) \not\leq aI \right\} \leq d \|E \exp t X_i\|^N \exp(-atN).$$

But $\exp tx \leq 1 + x(\exp t - 1)$ for $x \in [0, 1]$; hence

$$\exp t X_i \leq I + X_i(\exp t - 1),$$

and the right hand side is less than or equal to

$$d [1 + \mu(\exp t - 1)]^N \exp(-atN).$$

Taking $\exp t = \frac{a}{\mu} \frac{1-\mu}{1-a} \geq 1$ provides the right hand side of (10.78). Similarly, for $0 \leq a \leq \mu \leq 1$,

$$P \left\{ \omega : \frac{1}{N} \sum_{i=1}^N X_i(\omega) \not\geq aI \right\} \leq d \exp(-Nh(a; \mu)). \quad (10.79)$$

Note that $h(\mu; \mu) = 0$, $\frac{\partial}{\partial a} h(a; \mu)|_{\mu=a} = 0$, $\frac{\partial^2}{\partial^2 a} h(a; \mu) = \frac{1}{a(1-a)}$. It follows that

$$\frac{\partial^2}{\partial^2 \varepsilon} h(\mu(1 \pm \varepsilon); \mu) = \frac{\mu}{(1 \pm \varepsilon)(1 - \mu(1 \pm \varepsilon))} \geq \frac{\mu}{2},$$

if $\mu \leq \frac{1}{2}$, $|\varepsilon| \leq 1$. Hence,

$$h(\mu(1 \pm \varepsilon); \mu) \geq \frac{\mu \varepsilon^2}{4}. \quad (10.80)$$

Finally, let $X_i(\omega)$ satisfy the conditions of the Theorem 10.34. Consider the operator random variables $Y_i(\omega) = \mu M^{-1/2} X_i(\omega) M^{-1/2}$, such that $0 \leq Y_i(\omega) \leq I$ and $EY_i = \mu I$. We have

$$\begin{aligned} \left\{ \omega : \left\| \frac{1}{N} \sum_{i=1}^N X_i(\omega) - M \right\|_1 < \varepsilon \text{Tr } M \right\} \\ &\supseteq \left\{ \omega : (1 - \varepsilon)M \leq \frac{1}{N} \sum_{i=1}^N X_i(\omega) \leq (1 + \varepsilon)M \right\} \\ &= \left\{ \omega : (1 - \varepsilon)\mu \leq \frac{1}{N} \sum_{i=1}^N Y_i(\omega) \leq (1 + \varepsilon)\mu \right\} \\ &= \left\{ \omega : (1 - \varepsilon)\mu \leq \frac{1}{N} \sum_{i=1}^N Y_i(\omega) \right\} \\ &\quad \cap \left\{ \omega : \frac{1}{N} \sum_{i=1}^N Y_i(\omega) \leq (1 + \varepsilon)\mu \right\}. \end{aligned}$$

By taking the complements and applying the estimates (10.78) and (10.79), together with (10.80), we get (10.76). \square

10.4.4 The Direct Coding Theorem for the quantum capacity

The proof of inequality $Q(\Phi) \geq \overline{Q}(\Phi)$ will be based on a “coherent” version of the proof for the private classical capacity, i.e. a version that uses the same random coding, but where the mixtures of states are replaced with superpositions. It is sufficient to prove that for a given input state S the quantity $R = I_c(S, \Phi) - \delta$ is an achievable rate. Achievability of $\overline{Q}(\Phi) - \delta$ is then shown by applying the argument to $\Phi^{\otimes n}$ and taking the limit $n \rightarrow \infty$. In this section, we will show how to construct an encoding and decoding such that

$$F_e(\bar{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)}) \rightarrow 1, \quad (10.81)$$

where $\bar{S}^{(n)}$ is the chaotic state in $\mathcal{H}^{(n)}$, $d_n = \dim \mathcal{H}^{(n)} = 2^{nR}$. By the argument following Definition 10.19 this means that the rate R is achievable for asymptotically perfect transmission of quantum information.

For channel Φ , we consider representation (6.7) (Theorem 6.9), namely

$$\Phi[S] = \text{Tr}_E V S V^*,$$

where V is an isometric map from the input space \mathcal{H}_A to the tensor product $\mathcal{H}_B \otimes \mathcal{H}_E$ of the output and the environment spaces.

Consider a spectral decomposition $S = \sum_x \pi_x S_A^x$, where $S_A^x = |\phi_x\rangle_A \langle \phi_x|$, and $\{|\phi_x\rangle_A\}$ form an orthonormal basis. Denote by $|\phi'_x\rangle_{BE} = V|\phi_x\rangle_A$ the joint output states of BE . In this case, the states of B and E , correspondingly, will be $S_B^x = \text{Tr}_E |\phi'_x\rangle_{BE} \langle \phi'_x|$ and $S_E^x = \text{Tr}_B |\phi'_x\rangle_{BE} \langle \phi'_x|$. Let us recall that, according to (10.58),

$$I_c(S, \Phi) = \chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}). \quad (10.82)$$

Our aim is to construct a sequence of subspaces $\mathcal{H}^{(n)}$ such that $\dim \mathcal{H}^{(n)} = 2^{nR}$, $R = I_c(S, \Phi) - \delta$, and corresponding encodings and decodings for which (10.81) holds.

Now, consider the block channel $\Phi^{\otimes n}$ and the codebook

$$W^{(n)} = \{w^{mj}; m = 1, \dots, N_B; j = 1, \dots, N_E\},$$

used in the construction of the privacy code, with the corresponding resolution of the identity $M^{(n)} = \{M_{mj}\}$ in $\mathcal{H}_B^{\otimes n}$, such that

$$P_e(W^{(n)}, M^{(n)}) \equiv \max_{mj} (1 - \text{Tr } S_{B^{(n)}}^{mj} M_{mj}) < \varepsilon, \quad (10.83)$$

where $S_{B^{(n)}}^{mj} = S_B^{x_1} \otimes \cdots \otimes S_B^{x_n}$ if $w^{mj} = (x_1, \dots, x_n)$,

$$\left\| \frac{1}{N_E} \sum_{j=1}^{N_E} S_{E^{(n)}}^{mj} - \theta \right\|_1 < \sqrt{\varepsilon}; \quad m = 1, \dots, N_B. \quad (10.84)$$

Consider the space $\mathcal{H}^{(n)}$ with the orthonormal basis $\{|m\rangle; m = 1, \dots, N_B\}$, where

$$N_B = 2^n [\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}) - \delta] = 2^n [I_c(S, \Phi) - \delta],$$

and the encoding $\mathcal{E}^{(n)}$ given by the following isometric map from $\mathcal{H}^{(n)}$ to $\mathcal{H}_A^{\otimes n}$:

$$|m\rangle \rightarrow \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i\alpha_{mj}} |\phi_{mj}\rangle_A,$$

where $|\phi_{mj}\rangle_A = |\phi_{x_1}\rangle_A \otimes \cdots \otimes |\phi_{x_n}\rangle_A$ if $w^{mj} = (x_1, \dots, x_n)$, and $\{\alpha_{mj}\} = \alpha$ is a collection of phases to be chosen later. Isometry of this map follows from the fact that $\{|\phi_{mj}\rangle_A\}$ is an orthonormal system, because it is built from the orthonormal vectors $|\phi_x\rangle_A$.

This is what was called the coherent version of the private code for the classical information. The action of the encoding on the maximally entangled vector (10.38) produces the vector

$$\frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes \left[\frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i\alpha_{mj}} |\phi_{mj}\rangle_A \right].$$

The encoding followed by the channel $\Phi^{\otimes n}$ gives

$$\frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes \left[\frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i\alpha_{mj}} |\phi'_{mj}\rangle_{BE} \right], \quad (10.85)$$

where the vectors $|\phi'_{mj}\rangle_{BE} = V|\phi_{mj}\rangle_A$ are again orthonormal. Note that $S_{B^{(n)}}^{mj} = \text{Tr}_E |\phi'_{mj}\rangle_{BE} \langle \phi'_{mj}|$, $S_{E^{(n)}}^{mj} = \text{Tr}_B |\phi'_{mj}\rangle_{BE} \langle \phi'_{mj}|$, where, in order to simplify notations, we denote by Tr_E the partial trace with respect to $\mathcal{H}_E^{\otimes n}$ etc.

To construct the decoding $\mathcal{D}^{(n)}$, we first use the observable $M^{(n)} = \{M_{mj}\}$, which provides a measurement of mj with asymptotically vanishing error. By the discussion of the measurement process leading to relation (6.37), there are an auxiliary system $\mathcal{H}_0^{(n)} = \mathcal{H}^{(n)} \otimes \mathcal{H}_1^{(n)}$ with the basis $\{|m\rangle \otimes |j\rangle_1\}$, a unit vector $|\psi_0\rangle \in \mathcal{H}_0^{(n)}$, and a unitary operator U in $\mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_0^{(n)}$, such that

$$M_{mj} = \text{Tr}_0 (I_{B^{(n)}} \otimes |\psi_0\rangle \langle \psi_0|) U^* (I_{B^{(n)}} \otimes |mj\rangle_0 \langle mj|) U,$$

where we denoted $|mj\rangle_0 = |m\rangle \otimes |j\rangle_1$. Defining

$$|\psi_{mj}\rangle_{BE0} = (I_{E^{(n)}} \otimes U) \left(|\phi'_{mj}\rangle_{BE} \otimes |\psi_0\rangle \right),$$

we have the following state vector of the system $\mathcal{H}^{(n)} \otimes \mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_E^{\otimes n} \otimes \mathcal{H}_0^{(n)}$

$$|\Upsilon(\alpha)\rangle = \frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes \left[\frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i\alpha_{mj}} |\psi_{mj}\rangle_{BE0} \right],$$

as the result of the action of the encoding, followed by the channel, and further followed by the measurement, on the maximally entangled vector. Here, $\{|\psi_{mj}\rangle_{BE0}\}$ is again an orthonormal system. Relation (10.83) means that the “projection” of this system onto $\mathcal{H}_0^{(n)}$ is close to $\{|mj\rangle_0\}$. More precisely,

$${}_0\langle mj | (\text{Tr}_{BE} |\psi_{mj}\rangle_{BE0} \langle \psi_{mj}|) |mj\rangle_0 > 1 - \varepsilon \quad \text{for all } mj.$$

The following lemma implies that there exist unit vectors $|\chi_{mj}\rangle_{BE} \in \mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_E^{\otimes n}$ such that, for $|\tilde{\psi}_{mj}\rangle_{BE0} = |\chi_{mj}\rangle_{BE} \otimes |mj\rangle_0$, one has

$$|{}_{BE0}\langle \psi_{mj} | \tilde{\psi}_{mj} \rangle_{BE0}|^2 > 1 - \varepsilon. \quad (10.86)$$

Lemma 10.35. *Let $|\varphi\rangle_0 \in \mathcal{H}_0$, $|\psi\rangle \in \mathcal{H}_0 \otimes \mathcal{H}$ be two unit vectors. In this case,*

$${}_0\langle \varphi | (\text{Tr}_{\mathcal{H}} |\psi\rangle \langle \psi|) | \varphi \rangle_0 = \max_{\chi} |\langle \psi | (|\varphi\rangle_0 \otimes |\chi\rangle)|^2, \quad (10.87)$$

where the maximum is taken over all unit vectors $|\chi\rangle \in \mathcal{H}$.

Proof. By fixing an orthonormal basis $\{|e_j\rangle\}$ in \mathcal{H} , we have $|\chi\rangle = \sum_j c_j |e_j\rangle$, with $\sum_j |c_j|^2 = 1$ and

$$|\langle \psi | (|\varphi\rangle_0 \otimes |\chi\rangle)|^2 = \left| \sum_j c_j \langle \psi | (|\varphi\rangle_0 \otimes |e_j\rangle) \right|^2.$$

Maximizing with respect to c_j we obtain (10.87). □

Notice that $\{|\tilde{\psi}_{mj}\rangle_{BE0}\}$ is again an orthonormal system, approximating $\{|\psi_{mj}\rangle_{BE0}\}$ in the sense (10.86). Defining

$$|\Gamma(\alpha)\rangle = \frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes \left[\frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i\alpha_{mj}} |\tilde{\psi}_{mj}\rangle_{BE0} \right].$$

we have

$$\int \dots \int \langle \Upsilon(\alpha) | \Gamma(\alpha + \alpha') \rangle d\mu(\alpha) = \frac{1}{N_B N_E} \sum_{m=1}^{N_B} \sum_{j=1}^{N_E} e^{i\alpha'_{mj}} {}_{BE0} \langle \psi_{mj} | \tilde{\psi}_{mj} \rangle_{BE0},$$

where $d\mu(\alpha)$ is the uniform distribution. Here, the integration eliminates the inner products with $j \neq j'$. One can always pick $\alpha' = \{\alpha'_{mj}\}$ such that the right hand side will be

$$\frac{1}{N_B N_E} \sum_{m=1}^{N_B} \sum_{j=1}^{N_E} |{}_{BE0} \langle \psi_{mj} | \tilde{\psi}_{mj} \rangle_{BE0}| > \sqrt{1 - \varepsilon}$$

by (10.86). Hence, there is an α such that $\Re \langle \Upsilon(\alpha) | \Gamma(\alpha + \alpha') \rangle > \sqrt{1 - \varepsilon}$. By introducing an additional common phase to α'_{mj} we can always make the inner product positive. Hence,

$$\langle \Upsilon(\alpha) | \Gamma(\alpha + \alpha') \rangle > \sqrt{1 - \varepsilon}. \quad (10.88)$$

There is one more transformation to incorporate into the decoding $\mathcal{D}^{(n)}$, which will leave the system E in the state almost independent of m . To achieve this, note that combining (10.86) with (10.17) gives

$$\| |\psi_{mj}\rangle_{BE0} \langle \psi_{mj}| - |\tilde{\psi}_{mj}\rangle_{BE0} \langle \tilde{\psi}_{mj}| \|_1 < 2\sqrt{\varepsilon}. \quad (10.89)$$

Now note that $S_{E^{(n)}}^{mj} = \text{Tr } {}_{B0} |\psi_{mj}\rangle_{BE0} \langle \psi_{mj}|$ and denote

$$\tilde{S}_{E^{(n)}}^{mj} = \text{Tr } {}_{B0} |\tilde{\psi}_{mj}\rangle_{BE0} \langle \tilde{\psi}_{mj}| = \text{Tr } {}_B |\chi_{mj}\rangle_{BE} \langle \chi_{mj}|.$$

Then (10.89) and Exercise 10.25 imply

$$\| S_{E^{(n)}}^{mj} - \tilde{S}_{E^{(n)}}^{mj} \|_1 < 2\sqrt{\varepsilon}, \quad (10.90)$$

which, combined with (10.84), implies

$$\left\| \frac{1}{N_E} \sum_{j=1}^{N_E} \tilde{S}_{E^{(n)}}^{mj} - \theta \right\|_1 < O(\sqrt{\varepsilon}); \quad m = 1, \dots, N_B. \quad (10.91)$$

Consider the vectors

$$|\varphi_m\rangle_{BE1} = \frac{1}{\sqrt{N_E}} \sum_{j=1}^{N_E} e^{i(\alpha_{mj} + \alpha'_{mj})} |\chi_{mj}\rangle_{BE} \otimes |j\rangle_1,$$

which are purifications of $\frac{1}{N_E} \sum_{j=1}^{N_E} \bar{S}_{E^{(n)}}^{mj}$. Let $|\phi_\theta\rangle \in \mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_E^{\otimes n} \otimes \mathcal{H}_1^{(n)}$ be a purification of θ . Then (10.91) implies that there are unitary operators W_m in $\mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_1^{(n)}$ such that

$$\left| {}_{BE1} \langle \phi_\theta | \left(I_{\mathcal{H}_E^{\otimes n}} \otimes W_m \right) \varphi_m \rangle_{BE1} \right| > 1 - O(\sqrt{\varepsilon}) \quad \text{for all } m. \quad (10.92)$$

This follows from inequality (10.33), which says that if two states are close in trace norm, there exist purifications of these states with fidelity close to 1.

Defining a “controlled unitary” operator $W = \sum_{m=1}^{N_B} e^{i\beta_m} W_m \otimes |m\rangle\langle m|$ in $\mathcal{H}_B^{\otimes n} \otimes \mathcal{H}_0^{(n)}$, where β_m are phases to be adjusted later, we have

$$\left(I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) (|\varphi_m\rangle_{BE1} \otimes |m\rangle) = e^{i\beta_m} \left(I_{\mathcal{H}_E^{\otimes n}} \otimes W_m \right) |\varphi_m\rangle_{BE1} \otimes |m\rangle.$$

Hence,

$$\begin{aligned} & \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) |\Gamma(\alpha + \alpha')\rangle \\ &= \frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes |m\rangle \otimes e^{i\beta_m} \left(I_{\mathcal{H}_E^{\otimes n}} \otimes W_m \right) |\varphi_m\rangle_{BE1}. \end{aligned}$$

Denoting $|\Omega^{(n)}\rangle = \frac{1}{\sqrt{N_B}} \sum_{m=1}^{N_B} |m\rangle \otimes |m\rangle$ the maximally entangled vector in $\mathcal{H}^{(n)} \otimes \mathcal{H}^{(n)}$, we have

$$\begin{aligned} & \left(|\Omega^{(n)}\rangle \otimes \langle \phi_\theta| \right) \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) |\Gamma(\alpha + \alpha')\rangle \\ &= \frac{1}{N_B} \sum_{m=1}^{N_B} e^{i\beta_m} \langle \phi_\theta| \left(I_{\mathcal{H}_E^{\otimes n}} \otimes W_m \right) |\varphi_m\rangle_{BE1}. \end{aligned}$$

This can be made equal to

$$\frac{1}{N_B} \sum_{m=1}^{N_B} \left| \langle \phi_\theta| \left(I_{\mathcal{H}_E^{\otimes n}} \otimes W_m \right) |\varphi_m\rangle_{BE1} \right|$$

by an appropriate choice of phases β_m , which is greater than $1 - O(\sqrt{\varepsilon})$ by the estimate (10.92). Comparing this with (10.88), we have

$$\left(|\Omega^{(n)}\rangle \otimes \langle \phi_\theta| \right) \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) |\Upsilon(\alpha)\rangle > 1 - O(\sqrt{\varepsilon}),$$

as follows from the triangle inequality (10.31).

Introducing the density operator

$$S' = \text{Tr}_{BE1} \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) |\Upsilon(\alpha)\rangle \langle \Upsilon(\alpha)| \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right)$$

in $\mathcal{H}^{(n)} \otimes \mathcal{H}^{(n)}$, we have

$$\begin{aligned} F \left(|\Omega^{(n)}\rangle \langle \Omega^{(n)}|, S' \right) &\equiv \langle \Omega^{(n)} | S' | \Omega^{(n)} \rangle \\ &\geq \left| \left(\langle \Omega^{(n)} | \otimes \langle \phi_\theta | \right) \left(I_{\mathcal{H}^{(n)}} \otimes I_{\mathcal{H}_E^{\otimes n}} \otimes W \right) |\Upsilon(\alpha)\rangle \right|^2 \\ &> 1 - O(\sqrt{\varepsilon}). \end{aligned} \quad (10.93)$$

On the other hand,

$$S' = \left(\text{Id}_{\mathcal{H}^{(n)}} \otimes \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)} \right) \left[|\Omega^{(n)}\rangle \langle \Omega^{(n)}| \right],$$

where the decoding $\mathcal{D}^{(n)}$ is defined as

$$\mathcal{D}^{(n)} [S_{B^{(n)}}] = \text{Tr}_{B1} W U (S_{B^{(n)}} \otimes |\psi_0\rangle \langle \psi_0|) U^* W^*$$

(remember that $|\psi_0\rangle \in \mathcal{H}_0^{(n)} = \mathcal{H}^{(n)} \otimes \mathcal{H}_1^{(n)}$). Then (10.93) is the same as

$$F_e \left(\tilde{S}^{(n)}, \mathcal{D}^{(n)} \circ \Phi^{\otimes n} \circ \mathcal{E}^{(n)} \right) > 1 - O(\sqrt{\varepsilon}),$$

and (10.81) is established. \square

10.5 Notes and references

1. The first examples of quantum error-correcting codes were constructed by Shor [35] and by Steane [201]. Many authors contributed to the subsequent development of the subject, sketched in this chapter, see a survey in Nielsen and Chuang [158], and Gottesman [72]. Necessary and sufficient conditions for error correction of Theorem 10.3 were proposed by Knill and Laflamme [133]. For the solution of Exercise 10.4, see [158], n. 10.2. The possibility of error correction is of basic importance to the problem of the realization of a quantum computer, see, e.g. Nielsen and Chuang [158], and Valiev [210]. An architecture of a fault-tolerant quantum computer was proposed, correcting errors not only in the quantum register, but also in the error-correcting modules, see Kitaev [135] and Steane [201].

The importance of coherent information to perfect error correction was pointed out by Barnum, Nielsen, and Schumacher [16].

2. The relations between the fidelity measures were studied by Fuchs [62], Barnum, Nielsen, and Schumacher [16] and Werner and Kretschmann [140]. Fidelity for mixed states was investigated by Uhlmann [208]. Lemma 10.17 goes back to Powers and Störmer [169], see also [93].

3. The Coding Theorem for the quantum capacity was first conjectured by Lloyd [150] as $Q(\Phi) = \max_{\rho} I_c(\rho, \Phi)$, but it was soon recognized by Di Vincenzo, Shor, and Smolin [54] that \bar{Q}_n is strictly superadditive, see also Smith and Yard [200]. Hence, taking the limit in (10.43) is indeed required. Additional evidence in favor of a heuristic random subspace coding argument was given by Shor [191]. See also the discussion in Horodeckis' paper [117]. The proof based on this approach was given later by Hayden [81].

The inequality $Q(\Phi) \leq \bar{Q}(\Phi)$ (the Converse Coding Theorem), as well as lemmas in Section 10.3.1 were established by Barnum, Nielsen, and Schumacher [16] and Barnum, Knill, and Nielsen [15]. Corollary 10.22 is inspired by an observation of Smith and Smolin [198].

Relation (10.58) was obtained by Schumacher and Westmoreland [178], who used the ideas of classical information theory [3] to relate the “privacy” of information transmission via quantum channel Φ to the quantity

$$\chi(\{\pi_x\}, \{S_B^x\}) - \chi(\{\pi_x\}, \{S_E^x\}).$$

The notion of a degradable channel was introduced by Devetak and Shor [52], who showed that the quantum capacity of a degradable channel is given by the one-letter expression (10.53). This resembles the notion of a “stochastically degraded” channel [42] for classical broadcast channels, with the role of second receiver played by the environment. For such channels, there is a one-letter expression for the capacity region. Anti-degradable channels were considered by Caruso, Giovannetti, and Holevo [36]. The corollary 10.28 is obtained in the work of Cubitt, Ruskai and Smith [43].

Smith and Yard [200] provided an explicit example of a remarkable phenomenon named *superactivation*. There exist cases in which, given two quantum channels Φ_1, Φ_2 with zero quantum capacity, it is possible to have $Q(\Phi_1 \otimes \Phi_2) > 0$. The example in [200] is built by joining an anti-degradable channel Φ_1 with an entanglement-binding channel Φ_2 .

4. A complete proof of the Direct Coding Theorem ($Q(\Phi) \geq \bar{Q}(\Phi)$) was given by Devetak [51]. It is based on an idea that reveals a profound relation between the quantum and the private classical capacities, see (10.58). It is this proof (as well as a simpler proof of the converse) which we follow here. The large deviation estimates for random operators of the Bernstein type using the operator version (10.78) of Hoeffding's inequality ([89], Theorem 1) are due to Ahlswede and Winter [4]. Another proof, based on random subspaces, which is closer to initial argument of Lloyd, was given by Hayden, Shor, and Winter [84]. Proposition 10.31 is due to Smith [196]. For more recent results concerning tail bounds for sums of i.i.d. Hermitian matrices, including an improvement of the Ahlswede–Winter inequality using Lieb's theorem on convex trace functions in place of the Golden–Thompson inequality, see Tropp [207].

It was already stressed that a quantum channel is characterized by a whole spectrum of capacities. Apart from the capacities C , C_χ , C_{ea} , C_p , Q , treated in detail in this book, there are classical and quantum feedback capacities (denoted by C_b , Q_b correspondingly), and the classical and quantum capacities with a classical two-side communication (correspondingly C_2 , Q_2). In classical information theory, it is well known that feedback does not increase the Shannon capacity of a memoryless channel. In the quantum case, a similar property is established [29] for the entanglement-assisted capacity $C_{ea}(\Phi)$. Regarding the quantum capacity $Q(\Phi)$, it is known that it can not be increased by an additional unlimited forward classical communication [24, 16]. However, $Q(\Phi)$ can be increased if there is a possibility of transmitting the classical information in the backward direction. Such a protocol would allow one to create the maximum entanglement between the input and output, which can be used for quantum state teleportation. By this trick, even channels with zero quantum capacity supplemented with classical feedback can be used for the reliable transmission of quantum information [158, 29]. For a quantum channel, there is a hierarchy

$$\begin{array}{ccccccc} C_\chi & \leq & C_b & \leq & C_2 & \leq & C_{ea} \\ \vee | & & \vee | & & \vee | & & \vee | \\ Q & \leq & Q_b & \leq & Q_2 & \leq & Q_{ea} \end{array} ,$$

where \leq should be understood as “less than or equal to for all channels and strictly less for some channels”, see Bennett, Devetak, Shor, and Smolin [20]. Also, $C_{ea} = 2Q_{ea}$ and for some other pairs of capacities both inequalities are possible. Furthermore, there exists a so called “mother” protocol for information transmission that realizes all the protocols, by using additional resources (such as e.g. feedback or entanglement) [1]. For a very detailed and thoroughly systematized survey of the modern state of the art in the quantum Shannon theory, see the treatise of Wilde [221].

Part V

Infinite systems

Chapter 11

Channels with constrained inputs

The importance of quantum channels with constrained inputs was clear from the beginning of quantum communication. A main question motivating the emergence of quantum information theory was the capacity of an optical channel with constrained energy of the input signal. From a mathematical point of view, the necessity for constraints appears when the system describing the information carrier is infinite, which in the quantum case means using infinite-dimensional Hilbert space. One such class of systems, called “continuous variable” quantum systems, whose states are in a natural sense “Gaussian”, will be considered in the next chapter. One of the main goals of the present chapter is to obtain general expressions for the capacities of infinite-dimensional channels with constrained inputs, suitable for applications to quantum Gaussian channels with the energy constraint.

A new feature of the infinite-dimensional case is the discontinuity and unboundedness of the entropy of quantum states, which requires us to pay serious attention to the continuity of the entropic quantities. Another important feature of channels in infinite dimensions is the natural emergence of generalized “continuous” ensembles, understood as probability measures on the set of all quantum states. Various capacity-like quantities for a quantum channel involve optimization with respect to state ensembles that satisfy the appropriate constraint. The supremums in question turn out to be achievable under certain conditions, namely, continuity of the entropic quantity and the compactness of the constrained set. These two mathematical problems will be studied in present chapter. Another subject of this chapter is the structure of entanglement-breaking channels in infinite dimensions.

11.1 Convergence of density operators

In what follows, \mathcal{H} is a separable Hilbert space. A linear operator A , defined on \mathcal{H} , is called *bounded* if it maps the unit ball into a norm-bounded subset of \mathcal{H} . Most of the definitions in Chapter 2 carry over to bounded operators, and we shall explicitly mention when a modification should be made (see also [171], Ch. VI). In particular, the operator norm $\|A\|$ is defined as in (1.16) with the maximum replaced by the supremum. The operator A is bounded if and only if $\|A\| < \infty$. All bounded operators form the Banach algebra $\mathcal{B}(\mathcal{H})$. The trace norm $\|\cdot\|_1$ is defined as in (1.14), and the collection of all bounded operators with $\|T\|_1 < \infty$ forms the Banach space $\mathfrak{T}(\mathcal{H})$ of *trace-class* operators. The inequality

$$|\mathrm{Tr} TA| \leq \|T\|_1 \|A\| \quad (11.1)$$

holds for all $T \in \mathfrak{T}(\mathcal{H})$, $A \in \mathfrak{B}(\mathcal{H})$. The dual space of $\mathfrak{T}(\mathcal{H})$ is $\mathfrak{T}(\mathcal{H})^* = \mathfrak{B}(\mathcal{H})$, which means that every continuous linear functional on $\mathfrak{T}(\mathcal{H})$ has the form $T \rightarrow \text{Tr } TA$ for some $A \in \mathfrak{B}(\mathcal{H})$, with the norm equal to $\|A\|$. Notice also that

$$\|A\| \leq \|A\|_1 \quad (11.2)$$

and $\mathfrak{T}(\mathcal{H})$ is a proper subspace of $\mathfrak{B}(\mathcal{H})$ in the infinite-dimensional case.

A *density operator* (we also use the term *state*) is a positive trace-class operator with unit trace (cf. Definition 2.2). The state space $\mathfrak{S}(\mathcal{H})$ is the closed convex subset of $\mathfrak{T}(\mathcal{H})$ consisting of all density operators in \mathcal{H} . It is complete separable metric space with the metric defined by the trace norm: $\rho(S_1, S_2) = \|S_1 - S_2\|_1$.

A sequence of bounded operators $\{A_n\}$ in \mathcal{H} *weakly converges* to an operator A if $\lim_{n \rightarrow \infty} \langle \psi | A_n | \phi \rangle \rightarrow \langle \psi | A | \phi \rangle$ for all $\phi, \psi \in \mathcal{H}$. It turns out that the weak convergence, which is in general weaker than the trace norm convergence, coincides with it on $\mathfrak{S}(\mathcal{H})$:

Lemma 11.1. *Let $\{S_n\}$ be a sequence of density operators in \mathcal{H} weakly converging to a density operator S . In this case, it converges in the trace norm.*

Proof. For any finite-dimensional projector P ,

$$\begin{aligned} \|S_n - S\|_1 &\leq \|P(S_n - S)P\|_1 + 2\|PS_n(I - P)\|_1 + 2\|PS(I - P)\|_1 \\ &\quad + \|(I - P)S_n(I - P)\|_1 + \|(I - P)S(I - P)\|_1. \end{aligned}$$

The first term on the right tends to zero for any choice of P , since $PS_nP \rightarrow PSP$ due to the weak convergence, and due to the equivalence of all kinds of convergence in the finite-dimensional case. For the last two terms we have

$$\|(I - P)S(I - P)\|_1 = \text{Tr } (I - P)S(I - P) = \text{Tr } (I - P)S = 1 - \text{Tr } PS,$$

which can be made arbitrarily small by the choice of P , and

$$\|(I - P)S_n(I - P)\|_1 = 1 - \text{Tr } PS_n \rightarrow 1 - \text{Tr } PS$$

by the weak convergence. Finally, the intermediate terms can be evaluated as follows:

$$\|PS_n(I - P)\|_1 = \text{Tr } U^*PS_n(I - P) = \text{Tr } U^*P\sqrt{S_n}\sqrt{S_n}(I - P),$$

where U is the unitary operator from the polar decomposition of $PS_n(I - P)$. Now, by the operator Cauchy–Schwarz inequality for the trace

$$\text{Tr } U^*P\sqrt{S_n}\sqrt{S_n}(I - P) \leq \sqrt{\text{Tr } PS_n}\sqrt{1 - \text{Tr } PS_n} \rightarrow \sqrt{\text{Tr } PS}\sqrt{1 - \text{Tr } PS}$$

which can again be made small by the choice of P . □

In what follows, when speaking of convergence of density operators or states, we will have in mind the convergence in the sense of this lemma. A subset $K \subseteq \mathfrak{S}(\mathcal{H})$ is *compact* if any sequence $\{S_n\} \subset \mathfrak{S}(\mathcal{H})$ contains a subsequence that converges to a state $S \in K$.

Theorem 11.2. *A trace norm closed subset K of $\mathfrak{S}(\mathcal{H})$ is compact if and only if for arbitrary $\varepsilon > 0$ there exists a finite rank projector P_ε such that $\text{Tr } P_\varepsilon S > 1 - \varepsilon$ for all $S \in K$.*

Proof. Let K be a trace norm compact subset of $\mathfrak{S}(\mathcal{H})$. Suppose that there exists $\varepsilon > 0$ such that for an arbitrary finite rank projector P there exists a state $S \in K$ such that $\text{Tr } PS \leq 1 - \varepsilon$. Let P_n be a sequence of finite rank projectors in \mathcal{H} converging monotonously to the identity operator $I_{\mathcal{H}}$ in the weak operator topology and let S_n be the corresponding sequence of states in K . By the compactness of K , there exists a subsequence S_{n_k} that converges to a state $S_0 \in K$. By construction $\text{Tr } P_{n_l} S_{n_k} \leq \text{Tr } P_{n_k} S_{n_k} \leq 1 - \varepsilon$ for $k > l$. Hence,

$$\text{Tr } S_0 = \lim_{l \rightarrow +\infty} \text{Tr } P_{n_l} S_0 = \lim_{l \rightarrow +\infty} \lim_{k \rightarrow +\infty} \text{Tr } P_{n_l} S_{n_k} \leq 1 - \varepsilon,$$

which contradicts the fact that $S_0 \in K \subseteq \mathfrak{S}(\mathcal{H})$.

Conversely, let K be a closed subset of $\mathfrak{S}(\mathcal{H})$ that satisfies the criterion, and let S_n be an arbitrary sequence in K . Since the unit ball in $\mathfrak{B}(\mathcal{H})$ is compact in the weak operator topology (this follows from the Banach–Alaoglu Theorem, see e.g. [171], Theorem VI.21), there exists a subsequence S_{n_k} that converges weakly to a positive operator S_0 . We have

$$\text{Tr } S_0 \leq \liminf_{k \rightarrow \infty} \text{Tr } S_{n_k} = 1,$$

Therefore, to prove that S_0 is a state, it is sufficient to show that $\text{Tr } S_0 \geq 1$. Let $\varepsilon > 0$ and P_ε be the corresponding projector. We have

$$\text{Tr } S_0 \geq \text{Tr } P_\varepsilon S_0 = \lim_{k \rightarrow \infty} \text{Tr } P_\varepsilon S_{n_k} > 1 - \varepsilon,$$

where the equality follows from the fact that P_ε has finite rank. Thus, S_0 is a state. Lemma 11.1 implies that the subsequence S_{n_k} converges to the state S_0 in the trace norm. Thus, the set K is compact. \square

In the sequel we will need the following partial infinite-dimensional analog of the spectral decomposition (1.5). Let $\{|e_j\rangle\}$ be an orthonormal basis in \mathcal{H} and $\{f_j\}$ a sequence of real numbers bounded from below. In this case, the formula

$$F|\psi\rangle = \sum_j f_j |e_j\rangle \langle e_j | \psi \rangle \quad (11.3)$$

defines a self-adjoint operator F (see Section 12.1.1) on the dense domain

$$\mathcal{D}(F) = \left\{ \psi : \sum_j |f_j|^2 |\langle e_j | \psi \rangle|^2 < \infty \right\}, \quad (11.4)$$

for which $|e_j\rangle$ are the eigenvectors and f_j are the corresponding eigenvalues.

Definition 11.3. An operator defined on the domain (11.4) by the formula (11.3) will be called an *operator of the type \mathfrak{F}* .

In applications, F is the operator of energy (of an oscillator system). In connection with this, an important role will be played by the operator $\exp(-\theta F)$, $\theta > 0$, defined by the relation

$$\exp(-\theta F) |\psi\rangle = \sum_j \exp(-\theta f_j) |e_j\rangle \langle e_j | \psi \rangle, \quad (11.5)$$

which is a bounded positive operator of the type \mathfrak{F} .

Let F be an operator of the type \mathfrak{F} . For an arbitrary density operator S , we define the expectation

$$\text{Tr } SF = \sum_{j=1}^{\infty} f_j \langle e_j | S | e_j \rangle \leq +\infty \quad (11.6)$$

which is correctly defined, with values in $(-\infty, +\infty]$.

Lemma 11.4. The functional $S \rightarrow \text{Tr } SF$ is affine lower semicontinuous on the set $\mathfrak{S}(\mathcal{H})$.

Proof. Affinity is obvious. Next, we will use the fact that the least upper bound of a family of continuous functions is lower semicontinuous. We have

$$\text{Tr } SF = \sup_N \sum_{j=1}^N f_j \langle e_j | S | e_j \rangle,$$

where all the finite sums are continuous. Hence, $S_n \rightarrow S$ implies

$$\liminf_{n \rightarrow \infty} \text{Tr } S_n F \geq \text{Tr } SF,$$

which means the lower semicontinuity. □

Lemma 11.5. Let the spectrum of the operator F consist of the eigenvalues f_n of finite multiplicity and $\lim_{n \rightarrow \infty} f_n = +\infty$. In this case, the set

$$\mathfrak{S}_E = \{S : \text{Tr } SF \leq E\} \quad (11.7)$$

is compact.

Proof. Without loss of generality, we assume that f_n is monotonously nondecreasing, and denote by P_n the finite dimensional projector onto the eigenspace corresponding to the first n eigenvalues, so that $P_n \uparrow I$. The set \mathfrak{S}_E is closed, due to Lemma 11.4. Since $f_{n+1}(I - P_n) \leq F$, we have $\text{Tr } S(I - P_n) \leq f_{n+1}^{-1} \text{Tr } SF \leq f_{n+1}^{-1} E < \varepsilon$ for large enough n and for all $S \in \mathfrak{S}_E$. By the criterion of compactness of Theorem 11.2, \mathfrak{S}_E is compact. \square

In applications this result implies the compactness of the set of quantum states with mean energy bounded by a certain constant.

11.2 Quantum entropy and relative entropy

In the infinite-dimensional case, the von Neumann entropy $H(S)$ of a density operator S can be defined as in formula (5.7), although now it can take the value $+\infty$ if the corresponding series diverges. The relative entropy can be correctly defined by using formula (7.2). Most of the properties in Chapter 7 can be extended to this case, except for the continuity, which is weakened to the lower semicontinuity.

Theorem 11.6. *The quantum entropy and relative entropy are lower semicontinuous on $\mathfrak{S}(\mathcal{H})$. Let $\{S_n\}$ (resp. $\{S'_n\}$) be a sequence of density operators in \mathcal{H} , converging to a density operator S (resp. S'). In this case,*

$$H(S) \leq \liminf_{n \rightarrow \infty} H(S_n),$$

$$H(S; S') \leq \liminf_{n \rightarrow \infty} H(S_n; S'_n).$$

Proof. By inequality (11.2), we have $\|S_n - S\| \rightarrow 0$. The function $\eta(x) = -x \log x$ is continuous on the interval $[0, 1]$. Hence, $\|\eta(S_n) - \eta(S)\| \rightarrow 0$. Indeed, $\eta(x)$ can be uniformly approximated by a polynomial on $[0, 1]$, and then we can use the following estimate.

Exercise 11.7. For a polynomial f and arbitrary operators A, B of norm less than or equal to one

$$\|f(A) - f(B)\| \leq c_f \|A - B\|,$$

where the constant depends only on f . Hint: use the decomposition

$$A^k - B^k = \sum_{l=0}^{k-1} A^{k-l-1} (A - B) B^l.$$

Therefore, for any finite-dimensional projector P , due to inequality (11.1) we have

$$|\text{Tr } P(\eta(S_n) - \eta(S))| \leq \text{Tr } P \|\eta(S_n) - \eta(S)\| \rightarrow 0. \quad (11.8)$$

On the other hand, by the same inequality, for a Hermitian positive operator A ,

$$\mathrm{Tr} PA \leq \|P\| \mathrm{Tr} A$$

and hence $\mathrm{Tr} A = \sup_P \mathrm{Tr} PA$, where P runs over all finite-dimensional projectors. Thus,

$$H(S) = \sup_P \mathrm{Tr} P \eta(S) \leq \liminf_{n \rightarrow \infty} \sup_P \mathrm{Tr} P \eta(S_n) = \liminf_{n \rightarrow \infty} H(S_n).$$

Similarly, by using representation (7.23) for the relative entropy, we get

$$H(S; S') = \sup_{\lambda > 0} \frac{1}{\lambda} [H(\lambda S + (1 - \lambda)S') - \lambda H(S) - (1 - \lambda)H(S')],$$

hence

$$\begin{aligned} H(S; S') &= \sup_{P, \lambda > 0} \frac{1}{\lambda} \mathrm{Tr} P [\eta(\lambda S + (1 - \lambda)S') - \lambda \eta(S) - (1 - \lambda)\eta(S')] \\ &\leq \liminf_{n \rightarrow \infty} \sup_{P, \lambda > 0} \frac{1}{\lambda} \mathrm{Tr} P [\eta(\lambda S_n + (1 - \lambda)S'_n) - \lambda \eta(S_n) - (1 - \lambda)\eta(S'_n)] \\ &= \liminf_{n \rightarrow \infty} H(S_n; S'_n). \end{aligned}$$

□

Lemma 11.8. *Let F be an operator of type \mathfrak{F} , satisfying the condition*

$$\mathrm{Tr} \exp(-\theta F) < \infty \quad \text{for all } \theta > 0, \quad (11.9)$$

In this case, the quantum entropy $H(S)$ is bounded and continuous on the set \mathfrak{S}_E defined in (11.7).

Notice that relation (11.9) implies that the operator F satisfies the conditions of Lemma 11.5, and hence the set \mathfrak{S}_E is compact.

Proof. By introducing the density operator $S_\theta = \exp(-\theta F - c(\theta))$, where $c(\theta) = \ln \mathrm{Tr} \exp(-\theta F)$, we have

$$H(S) = -H(S; S_\theta) + \theta \mathrm{Tr} SF + c(\theta), \quad (11.10)$$

where $H(S; S_\theta)$ is the relative entropy. Hence,

$$H(S) \leq -H(S; S_\theta) + \theta E + c(\theta), \quad (11.11)$$

if $S \in \mathfrak{S}_E$. Therefore, the entropy $H(S)$ is bounded on \mathfrak{S}_E . Since $H(S)$, by Theorem 11.6, is lower semicontinuous, it is sufficient to show that it is also upper semicontinuous, and hence continuous, on the compact set \mathfrak{S}_E . Let $\{S_n\} \subset \mathfrak{S}_E$ be a

sequence of states weakly converging to S . From inequality (11.11) applied to S_n , by lower semicontinuity of the relative entropy,

$$\begin{aligned} \limsup_{n \rightarrow \infty} H(S_n) &\leq -H(S; S_\theta) + \theta E + c(\theta) \\ &= H(S) + \theta(E - \text{Tr } SF), \end{aligned} \quad (11.12)$$

where in the last equality we used (11.10). Letting $\theta \rightarrow 0$, we obtain the upper semicontinuity of $H(S)$. \square

Notice that when F is the operator of the energy, S_θ is the density operator of the Gibbs equilibrium state at the inverse temperature θ , and the function $-\theta^{-1}c(\theta)$ is the free energy. The relation (11.10) and the non-negativity of the relative entropy imply

$$H(S) \leq \theta \text{Tr } SF + c(\theta), \quad (11.13)$$

which is equivalent to the Gibbs variational principle.

11.3 Constrained c-q channel

Let $\mathcal{X} = \{x\}$ be an infinite alphabet. For every x , let S_x be a density operator in \mathcal{H} with finite von Neumann entropy $H(S_x)$. The map $x \rightarrow S_x$ will be called a *c-q channel* with input alphabet \mathcal{X} . Let $f(x)$ be a function defined on \mathcal{X} , taking values in $[0, +\infty]$. We shall consider the class \mathcal{P}_E of finitely supported probability distributions $\pi = \{\pi_x\}$ on \mathcal{X} satisfying the condition

$$\sum_x f(x)\pi_x \leq E, \quad (11.14)$$

where E is a positive real number. We assume that \mathcal{P}_E is nonempty and impose the following condition onto the channel:

$$\sup_{\pi \in \mathcal{P}_E} H\left(\sum_x \pi_x S_x\right) < \infty. \quad (11.15)$$

Definition 11.9. We define a *code* (W, M) of size N and length n as in Definition 5.1, with the additional requirement that all codewords $w = (x_1, \dots, x_n) \in W$ satisfy the additive constraint

$$f(x_1) + \dots + f(x_n) \leq nE, \quad (11.16)$$

The average error $\tilde{P}_e(W, M)$ and the minimal error $\bar{p}_e(n, N)$ of the code are given by formulas (5.5) and (5.6), respectively. The *constrained classical capacity* of channel $x \rightarrow S_x$ is defined as in Definition 5.2, i.e. as the supremum of achievable rates R such that $\lim_{n \rightarrow \infty} \bar{p}_e(n, 2^{nR}) = 0$.

Theorem 11.10. *The constrained classical capacity C of a c - q channel $x \rightarrow S_x$ that satisfies condition (11.15), with input constraint (11.16) is equal to the quantity*

$$C_\chi = \sup_{\pi \in \mathcal{P}_E} \left[H \left(\sum_x \pi_x S_x \right) - \sum_x \pi_x H(S_x) \right]. \quad (11.17)$$

Proof. We denote by $\mathcal{P}_E^{(n)}$ the class of probability distributions on \mathcal{X}^n satisfying the condition

$$\sum_{x_1, \dots, x_n} [f(x_1) + \dots + f(x_n)] \pi_{x_1, \dots, x_n} \leq nE \quad (11.18)$$

and define the quantity $C_\chi^{(n)}$ as in (5.30), with the modification that the supremum with respect to π is taken over $\mathcal{P}_E^{(n)}$, i.e.

$$C_\chi^{(n)} = \sup_{\pi \in \mathcal{P}_E^{(n)}} \chi(\{\pi_w\}; \{S_w\}).$$

Lemma 11.11. *The sequence $\{C_\chi^{(n)}\}$ is additive, i.e. $C_\chi^{(n)} = nC_\chi$.*

Proof. By (5.31), we have

$$\chi_n(\pi) \leq \sum_{k=1}^n \chi(\pi^{(k)}),$$

where $\pi^{(k)}$ is the k -th marginal distribution of π on \mathcal{X} . Also,

$$\sum_{k=1}^n \chi(\pi^{(k)}) \leq n\chi(\bar{\pi}), \quad (11.19)$$

where $\bar{\pi} = \frac{1}{n} \sum_{k=1}^n \pi^{(k)}$, since $\chi(\pi)$ is a concave function of π (as follows from concavity of the von Neumann entropy). Inequality (11.18) can be rewritten as

$$\frac{1}{n} \sum_{k=1}^n \sum_x f(x) \pi^{(k)}(x) \leq E,$$

which implies that $\bar{\pi} \in \mathcal{P}_E$ if $\pi \in \mathcal{P}_E^{(n)}$. Taking the supremum in relation (11.19) with respect to $\pi \in \mathcal{P}_E^{(n)}$, we obtain $C_\chi^{(n)} \leq nC_\chi$. The converse inequality is obvious. \square

The proof of the Converse Coding Theorem can be based on the following corollary of the Fano inequality (cf. (5.33)),

$$\overline{P}_e(W, M) \geq 1 - \frac{\sup_{\pi \in \mathcal{P}_E^{(n)}} \sup_M J_n(\pi, M)}{nR} - \frac{1}{nR}, \quad (11.20)$$

which is obtained as follows. Let again the words in the code (W, M) be taken with the input distribution $\pi^{(N)}$ assigning equal probability $1/N$ to each word. Consider inequality (4.37). Since the words in the code satisfy (11.16), we have $\pi^{(N)} \in \mathcal{P}_E^{(n)}$, and hence (11.20) follows.

Using inequality (11.20) and Theorem 5.9, we obtain

$$\bar{p}_e(n, 2^{nR}) \geq 1 - \frac{C_\chi}{R} - \frac{1}{nR},$$

where C_χ is defined by (11.17), which implies $p_e(n, 2^{nR}) \not\rightarrow 0$ for $R > C_\chi$.

In the classical information theory, the Direct Coding Theorem for channels with additive constraints can be proved by using random coding with a probability distribution (5.47) modified by a factor concentrated on words, for which the constraint holds close to the equality. The same tool can be applied to a c-q channel. Let π be a distribution on \mathcal{X} satisfying (11.14), and let P be a distribution on the set of N words, under which the words are independent and have the probability distribution (5.47). Let $v_n = P(\frac{1}{n} \sum_{k=1}^n f(x_k) \leq E)$ and define the modified distribution \tilde{P} under which the words are still independent, but

$$\tilde{P}(w = (x_1, \dots, x_n)) = \begin{cases} v_n^{-1} \pi_{x_1} \cdots \pi_{x_n}, & \text{if } \sum_{k=1}^n f(x_k) \leq nE, \\ 0, & \text{otherwise.} \end{cases} \quad (11.21)$$

Let us remark that since $\pi \in \mathcal{P}_E$, then $E f \leq E$ (where E is the expectation corresponding to P) and hence, by the Central Limit Theorem,

$$\lim_{n \rightarrow \infty} v_n \geq 1/2.$$

Therefore, $\tilde{E}\xi \leq 2^m E\xi$ (where \tilde{E} is the expectation corresponding to \tilde{P}) for any non-negative random variable ξ depending on m different words.

For the error probability $\bar{P}_e(W, M)$, we have the basic upper bound (5.52). Now, take the expectation of this bound with respect to \tilde{P} . Since every term in the right hand side of (5.52) depends on no more than two different words, we have

$$\tilde{E} \inf_M \bar{P}_e(W, M) \leq 4E \inf_M \bar{P}_e(W, M),$$

and the expectation with respect to P can be made arbitrarily small for $N = 2^{nR}$, $n \rightarrow \infty$, with $R < C_\chi - 3\delta$. Thus, $\tilde{E} \bar{P}_e(W, M)$ can also be made arbitrarily small under the same circumstances. Since the distribution \tilde{P} is concentrated on words that satisfy (11.16), we can choose a code for which $\bar{P}_e(W, M)$ can be made arbitrarily small for sufficiently large n . \square

11.4 Classical-quantum channel with continuous alphabet

Let the input alphabet \mathcal{X} be a complete, separable metric space with the σ -algebra of Borel subsets. In applications, \mathcal{X} is typically a domain in \mathbb{R}^k that is locally compact. However, we need the more general case for the treatment of infinite-dimensional quantum channels in the next section. We consider the c-q channel given by a *continuous* mapping $x \rightarrow S_x$ from the alphabet \mathcal{X} to the set of quantum states $\mathfrak{S}(\mathcal{H})$ (by Lemma 11.1, for continuity it is necessary and sufficient that all matrix elements $\langle \psi | S_x | \phi \rangle$; $\psi, \phi \in \mathcal{H}$ are continuous with respect to the metric in \mathcal{X}). We will obtain a convenient integral expression for the classical capacity C of the channel $x \rightarrow S_x$ under the constraint (11.16).

We assume that f is a nonnegative Borel function on \mathcal{X} . Consider the set \mathcal{P}_E^B of Borel probability measures π on \mathcal{X} satisfying

$$\int_{\mathcal{X}} f(x) \pi(dx) \leq E. \quad (11.22)$$

Recall that a sequence of Borel probability measures $\{\pi_n\}$ *converges weakly* to π if

$$\int_{\mathcal{X}} \varphi(x) \pi_n(dx) \rightarrow \int_{\mathcal{X}} \varphi(x) \pi(dx)$$

for arbitrary $\varphi \in C(\mathcal{X})$, the Banach space of bounded continuous functions on \mathcal{X} (see e.g. Parthasarathy's book [164]).

Exercise 11.12. Let f be lower semicontinuous. In this case, the functional $\pi \rightarrow \int_{\mathcal{X}} f(x) \pi(dx)$ is lower semicontinuous with respect to the weak convergence of probability measures.

Hint: use the fact that f is the least upper bound of the family of all bounded continuous functions $\varphi \leq f$.

We will need the following auxiliary result, the proof of which can be found in [164]:

Lemma 11.13. *Let f be lower semicontinuous. In this case, the set \mathcal{P}_E of all finitely supported probability measures that satisfy (11.22) is weakly dense in \mathcal{P}_E^B .*

Let us also give a convenient sufficient condition of weak compactness of the set \mathcal{P}_E^B .

Lemma 11.14. *Assume that f is lower semicontinuous and for any positive k the set $\{x : f(x) \leq k\} \subset \mathcal{X}$ is compact. In this case, the subset of probability measures*

$$\mathcal{P}_E^B = \left\{ \pi : \int_{\mathcal{X}} f(x) \pi(dx) \leq E \right\}$$

is weakly compact.

Proof. The lower semicontinuity of f implies that the functional $\pi \rightarrow \int_{\mathcal{X}} f(x)\pi(dx)$ is lower semicontinuous with respect to the weak convergence of probability measures (Exercise 11.12). Hence, the set \mathcal{P}_E^B is weakly closed. It is known that a weakly closed subset \mathcal{P} of Borel probability measures on \mathcal{X} is weakly compact if and only if for any $\varepsilon > 0$ there is a compact $K \subset \mathcal{X}$, such that $\pi(\bar{K}) \leq \varepsilon$ for all $\pi \in \mathcal{P}$ (see e.g. [164]). For a given $\varepsilon > 0$, consider the compact set $K = \{x : f(x) \leq E/\varepsilon\}$. Then,

$$\pi(\bar{K}) \leq \frac{\varepsilon}{E} \int_{\bar{K}} f(x)\pi(dx) \leq \varepsilon$$

for $\pi \in \mathcal{P}_E^B$. Hence, the set \mathcal{P}_E^B is weakly compact. \square

For an arbitrary Borel probability measure π on \mathcal{X} , we introduce the average state

$$\bar{S}_\pi = \int_{\mathcal{X}} S_x \pi(dx), \quad (11.23)$$

where, by assumption, the function S_x is continuous, the integral is well defined, and represents a density operator in \mathcal{H} .

Assuming $H(\bar{S}_\pi) < \infty$, consider the functional

$$\chi(\pi) = H(\bar{S}_\pi) - \int_{\mathcal{X}} H(S_x)\pi(dx), \quad (11.24)$$

where, by Theorem 11.6, the function $H(S_x)$ is nonnegative and lower semicontinuous and hence the integral is well-defined.

Theorem 11.15. *Let there exist a self-adjoint operator F of the type \mathfrak{F} , satisfying condition (11.9), such that*

$$f(x) \geq \text{Tr } S_x F; \quad x \in \mathcal{X}. \quad (11.25)$$

In this case, condition (11.15) holds and the classical capacity C of the channel $x \rightarrow S_x$ with the constraint (11.16) is finite, and is given by the relation (11.17).

If, moreover, the function f satisfies the conditions of Lemma 11.14, then

$$C = \max_{\pi \in \mathcal{P}_E^B} \chi(\pi). \quad (11.26)$$

Proof. By integrating (11.25), we obtain

$$\text{Tr } \bar{S}_\pi F \leq E \quad (11.27)$$

for $\pi \in \mathcal{P}_E$. Hence, by (11.13),

$$H(\bar{S}_\pi) \leq \theta \text{Tr } \bar{S}_\pi F + c(\theta) \leq \theta E + c(\theta),$$

Therefore, condition (11.15) is fulfilled, C is finite and equal to (11.17).

If the function f satisfies the conditions of Lemma 11.14, the set \mathcal{P}_E^B is weakly compact. Let us show that the functional $\pi \rightarrow \chi(\pi)$ is upper semicontinuous. From expression (11.17) and Lemma 11.13 it now follows that

$$C = \sup_{\pi \in \mathcal{P}_E} \chi(\pi) = \sup_{\pi \in \mathcal{P}_E^B} \chi(\pi),$$

Moreover, the last supremum is achieved by upper semicontinuity of $\chi(\pi)$ and compactness of \mathcal{P}_E^B .

Consider the first term in expression (11.24) for $\chi(\pi)$. Since the matrix elements of the family $\{S_x\}$ depend continuously on x , the mapping $\pi \rightarrow \tilde{S}_\pi$ into the set \mathfrak{S}_E , equipped with the weak operator topology, is continuous with respect to the weak convergence of probability measures. By Lemma 11.1, the weak operator topology on $\mathfrak{S}(\mathcal{H})$ is equivalent to the trace norm topology. By Lemma 11.8, the entropy is continuous on \mathfrak{S}_E . Therefore, the functional $\pi \rightarrow H(\tilde{S}_\pi)$ is continuous on the set \mathcal{P}_E^B .

The function $x \rightarrow H(S_x)$ is lower semicontinuous. Hence, by Exercise 11.12, the second term in (11.24) is upper semicontinuous in π . Thus, the functional (11.24) is also upper semicontinuous. \square

11.5 Constrained quantum channel

Let $\mathcal{H}_A, \mathcal{H}_B$ be separable Hilbert spaces called the input and output spaces, respectively.

Definition 11.16. By *channel* we call a linear, bounded, trace-preserving, completely positive map $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$. As in the finite dimensional case, complete positivity means that, for any $n = 1, 2, \dots$, the map $\Phi \otimes \text{Id}_n$ is positive.

Exercise 11.17. Any affine map $\Phi_s : \mathfrak{S}(\mathcal{H}_A) \rightarrow \mathfrak{S}(\mathcal{H}_B)$ extends uniquely to a linear, bounded, positive map $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$. Hint: to define the linear extension, use the construction of Exercise 2.7. To prove boundedness, use the argument from the proof of Lemma 10.12 (estimates of the type (10.22), (10.23) for $\|\Phi[T]\|_1$).

Example 11.18. Let $x \rightarrow S_x$ be a c-q channel with the input alphabet \mathcal{X} , which is a complete separable metric space. Then it can be extended to a quantum channel in the sense of Definition 11.16 as follows. Consider the Hilbert space $\mathcal{H}_A = L^2(\mathcal{X}, m)$, where m is a σ -finite measure on \mathcal{X} . Assume that S_x is an m -measurable family of density operators in \mathcal{H}_B . Any operator $S \in \mathfrak{T}(\mathcal{H}_A)$ is given by a kernel with a

correctly defined diagonal value $\langle x|S|x\rangle$. Then the relation

$$\Phi[S] = \int_{\mathcal{X}} \langle x|S|x\rangle S_x m(dx)$$

defines a channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$.

In this section, we study the classical capacity of a channel under the additive constraint at its input. Let F be an operator of type \mathfrak{F} in \mathcal{H}_A , representing an observable, the mean value of which is to be constrained. We assume that the input states $S^{(n)}$ of the composite channel $\Phi^{\otimes n}$ are subject to the additive constraint

$$\mathrm{Tr} S^{(n)} F^{(n)} \leq nE, \quad (11.28)$$

where

$$F^{(n)} = F \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes F,$$

and E is a positive constant. Adapting Definition 8.1 and the proof of Proposition 8.2 to the case of the input states constrained by (11.28), one can prove

Proposition 11.19. *Let the channel Φ satisfy the condition*

$$\sup_{S: \mathrm{Tr} SF \leq E} H(\Phi[S]) < \infty. \quad (11.29)$$

Then the classical capacity of this channel, under the constraint (11.28), is finite and is equal to

$$C(\Phi, F, E) = \lim_{n \rightarrow \infty} \frac{1}{n} C_{\chi}(\Phi^{\otimes n}, F^{(n)}, nE), \quad (11.30)$$

where

$$C_{\chi}(\Phi, F, E) = \sup_{\pi: \mathrm{Tr} \bar{S}_{\pi} F \leq E} \left[H(\Phi[\bar{S}_{\pi}]) - \sum_i \pi_i H(\Phi[S_i]) \right]. \quad (11.31)$$

Here, $\bar{S}_{\pi} = \sum_i \pi_i S_i$ is the average state of the ensemble $\pi = \{\pi_i, S_i\}$.

The finiteness of the capacity follows from

Lemma 11.20. *Condition (11.29) for the channel Φ implies a similar condition for the channel $\Phi^{\otimes n}$, namely*

$$\sup_{S^{(n)}: \mathrm{Tr} S^{(n)} F^{(n)} \leq nE} H(\Phi^{\otimes n}[S^{(n)}]) \leq n \sup_{S: \mathrm{Tr} SF \leq E} H(\Phi[S]). \quad (11.32)$$

Proof. Denoting by S_j the partial state of $S^{(n)}$ in the j -th tensor factor of $\mathcal{H}^{\otimes n}$, and letting $\bar{S} = \frac{1}{n} \sum_{j=1}^n S_j$, we have

$$H\left(\Phi^{\otimes n}[S^{(n)}]\right) \leq \sum_{j=1}^n H\left(\Phi[S_j]\right) \leq nH\left(\Phi[\bar{S}]\right),$$

where in the first inequality we use subadditivity of the quantum entropy, while in the second we use its concavity. Moreover, $\text{Tr } \bar{S} F = \frac{1}{n} \text{Tr } S^{(n)} F^{(n)} \leq E$ and hence (11.32) follows. \square

If the channel Φ satisfies the additivity property

$$C_\chi(\Phi^{\otimes n}, F^{(n)}, nE) = nC_\chi(\Phi, F, E), \quad (11.33)$$

then

$$C(\Phi, F, E) = C_\chi(\Phi, F, E).$$

This is closely related to the property of superadditivity of the convex hull of the output entropy (8.36), which implies the additivity of χ -capacity under the linear constraints (8.39) (see Section 8.3.2).

Exercise 11.21. Prove the following inequality

$$C_\chi(\Phi^{\otimes n}, F^{(n)}, nE) \geq nC_\chi(\Phi, F, E). \quad (11.34)$$

In any case, the quantity (11.31) provides a lower bound for the classical capacity $C(\Phi, F, E)$. We shall obtain a more convenient expression for $C_\chi(\Phi, F, E)$, by using the results of the previous section. Consider a c-q channel with the alphabet $\mathcal{X} = \mathfrak{S}(\mathcal{H}_A)$, defined by the mapping $S \rightarrow \Phi[S]$. The constraint is given by the function $f(S) = \text{Tr } SF$, which is affine and lower semicontinuous by Lemma 11.4, whereas the condition (11.15) turns into (11.29).

Definition 11.22. We call *generalized ensemble* an arbitrary Borel probability measure π on $\mathfrak{S}(\mathcal{H}_A)$. The *average state* of the generalized ensemble π is defined as the barycenter of the probability measure

$$\bar{S}_\pi = \int_{\mathfrak{S}(\mathcal{H}_A)} S \pi(dS).$$

The conventional ensembles correspond to finitely supported measures.

Theorem 11.6 implies, in particular, that the nonnegative function $S \rightarrow H(\Phi[S])$ is measurable. Hence, under condition $H(\Phi(\bar{S}_\pi)) < \infty$, the functional

$$\chi_\Phi(\pi) = H(\Phi(\bar{S}_\pi)) - \int_{\mathfrak{S}(\mathcal{H}_A)} H(\Phi(S))\pi(dS) \quad (11.35)$$

is well defined.

Exercise 11.23. The quantity $\chi_\Phi(\pi)$ is a concave functional of π . Hint: The second term is affine, the concavity of the first term follows from the concavity of the entropy.

Corollary 11.24. *Let there exist a positive operator F' in \mathcal{H}_B , such that*

$$\text{Tr} \exp(-\theta F') < +\infty, \quad \text{for all } \theta > 0 \quad (11.36)$$

and

$$\text{Tr} \Phi[S]F' \leq \text{Tr} SF; \quad S \in \mathfrak{S}(\mathcal{H}). \quad (11.37)$$

In this case, condition (11.29) is fulfilled and

$$C_\chi(\Phi, F, E) = \sup_{\pi: \text{Tr} \bar{S}_\pi F \leq E} \chi_\Phi(\pi).$$

Moreover, the output entropy $H(\Phi[S])$ is continuous on the compact set $\mathfrak{S}_E = \{S : \text{Tr} SF \leq E\}$.

If, additionally, F satisfies the conditions of Lemma 11.5, then there exists an optimal generalized ensemble, i.e.

$$C_\chi(\Phi, F, E) = \max_{\pi: \text{Tr} \bar{S}_\pi F \leq E} \chi_\Phi(\pi). \quad (11.38)$$

Proof. The statement is obtained by applying Theorem 11.15 to the c-q channel $S \rightarrow \Phi[S]$, with the constraint function $f(S) = \text{Tr} SF$, the role of condition (11.25) is played by (11.37), and the operator F' plays the role of F . The function $f(S) = \text{Tr} SF$ satisfies the conditions of Lemma 11.14, due to Lemmas 11.4, 11.5.

The continuity of the entropy $H(\Phi[S])$ follows from the continuity of the map $S \rightarrow S' = \Phi[S]$ and Lemma 11.8, which guarantees continuity of $H(S')$ on the compact set $\{S' : \text{Tr} S'F' \leq E\}$. \square

11.6 Entanglement-assisted capacity of constrained channels

Let systems A and B share an entangled (pure) state S_{AB} . We assume that the amount of entanglement is unlimited but finite, i.e. $H(S_A) = H(S_B) < \infty$. By generalizing the result of Section 9, one can prove

Proposition 11.25. *Let Φ be a channel that satisfies the condition (11.29) with the operator F satisfying (11.9). In this case, its entanglement-assisted classical capacity under the constraint (11.28) is finite and equals*

$$C_{ea}(\Phi) = \sup_{S: \text{Tr } SF \leq E} I(S, \Phi), \quad (11.39)$$

where

$$I(S, \Phi) = H(S) + H(\Phi[S]) - H(S; \Phi), \quad (11.40)$$

and $H(S; \Phi)$ is the entropy exchange.

Now, we investigate the problem when the supremum in the right hand side of (11.39) is achieved. Note that condition (11.9) implies that F satisfies the condition of the Lemma 11.5 and, hence, the set $\mathfrak{S}_E = \{S : \text{Tr } SF \leq E\}$ is compact.

Proposition 11.26. *Let the constraint operator F satisfy the condition (11.9), and let there exist a self-adjoint operator F' of the type \mathfrak{F} , satisfying (11.9) and (11.37). In this case,*

$$C_{ea}(\Phi) = \max_{S: \text{Tr } SF \leq E} I(S, \Phi). \quad (11.41)$$

Moreover, if the channel Φ is such that

$$\sup_S I(S, \Phi) = \infty, \quad (11.42)$$

the maximum in (11.41) is achieved on a density operator S , that satisfies the constraint with the equality $\text{Tr } SF = E$.

Proof. We will consider each term in formula (11.40) separately. Notice that, by Theorem 11.6, the quantum entropy is lower semicontinuous. Since the entropy exchange can be represented as $H(S; \Phi) = H(\tilde{\Phi}[S])$, where $\tilde{\Phi}$ is the complementary channel from the system space \mathcal{H}_A to the environment space \mathcal{H}_E , it is also lower semicontinuous and thus the last term in (11.40) is upper semicontinuous. Concerning the first term, by Lemma 11.8, it is continuous on the set $\mathfrak{S}_E = \{S : \text{Tr } SF \leq E\}$ if the constraint operator F satisfies (11.9). By using the proof of Corollary 11.24, we can apply a similar argument to the second term in (11.40), namely $H(\Phi[S])$. Moreover, this corollary also implies that condition (11.29) and hence (11.39) hold. As follows from the above, the mutual information (11.40) is upper semicontinuous on the compact set \mathfrak{S}_E , and hence attains its maximum.

To prove the second statement, we consider

$$f_0(E) = \max_{\text{Tr } SF = E} I(S, \Phi),$$

and assume that there exists a state S_1 such that

$$\text{Tr } S_1 F < E, \quad I(S_1, \Phi) > f_0(E).$$

Condition (11.42) implies that there exists a state S_2 such that

$$\text{Tr } S_2 F > E, \quad I(S_2, \Phi) > f_0(E).$$

Now, putting

$$\lambda = \frac{\text{Tr } S_2 F - E}{\text{Tr } S_2 F - \text{Tr } S_1 F},$$

we have $0 < \lambda < 1$ and $\text{Tr } (\lambda S_1 + (1 - \lambda) S_2) F = E$, so that

$$I(\lambda S_1 + (1 - \lambda) S_2, \Phi) \leq f_0(E) < \lambda I(S_1, \Phi) + (1 - \lambda) I(S_2, \Phi),$$

which contradicts the concavity of $I(S, \Phi)$. □

11.7 Entanglement-breaking channels in infinite dimensions

A state $S \in \mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is called *separable (unentangled)* if it belongs to the convex closure (i.e. closure of the convex hull) of the set of all product states $S_1 \otimes S_2 \in \mathfrak{S}(\mathcal{H}_1 \otimes \mathcal{H}_2)$.

Proposition 11.27. *State S is separable if and only if it admits a representation*

$$S = \int_{\mathcal{X}} S_1(x) \otimes S_2(x) \pi(dx), \quad (11.43)$$

where $\pi(dx)$ is the Borel probability measure and $S_j(x)$, $j = 1, 2$, are Borel $\mathfrak{S}(\mathcal{H}_j)$ -valued functions on some complete, separable metric space \mathcal{X} .

Proof. According to the definition, the state S is separable if and only if $S = \lim_{n \rightarrow \infty} S_n$, where

$$S_n = \int_{\mathfrak{S}(\mathcal{H}_1)} \int_{\mathfrak{S}(\mathcal{H}_2)} S_1 \otimes S_2 \pi_n(dS_1 dS_2), \quad (11.44)$$

and $\{\pi_n\}$ is a sequence of Borel probability measures with finite supports.

Let S be representable in the form (11.43). In this case, by making a change of variable $x \rightarrow S_1(x) \otimes S_2(x)$, we can reduce the integral representation (11.43) to

$$S = \int_{\mathfrak{S}(\mathcal{H}_1)} \int_{\mathfrak{S}(\mathcal{H}_2)} S_1 \otimes S_2 \pi(dS_1 dS_2), \quad (11.45)$$

where we use the same notation π for the image of the initial measure under the change of variable. The mapping $\pi \rightarrow S$ is continuous under weak convergence of probability measures. This is clear if the set of density operators is equipped with the weak operator topology and, by Lemma 11.1, this topology coincides with the

trace norm topology on this set. By Lemma 11.13, there is a sequence $\{\pi_n\}$ of Borel probability measures with finite supports, which converges weakly to π . By using the continuity of the mapping $\pi \rightarrow S$, we obtain $S = \lim_{n \rightarrow \infty} S_n$, where S_n is given by relation (11.44). Hence, S is separable.

Conversely, let S be a separable state. In this case, $S = \lim_{n \rightarrow \infty} S_n$ where S_n is given by (11.44). If we can prove that the sequence of measures $\{\pi_n\}$ is weakly relatively compact, this will imply representation (11.45), where π is a partial limit of this sequence. The converging sequence $\{S_n\}$ is relatively compact. Therefore, by Theorem 11.2, for any $\varepsilon > 0$ there is a finite-dimensional projection P such that $\text{Tr } S_n(I - P) \leq \varepsilon$ for all n . For $m = 1, 2, \dots$ denote by P_m the projection such that $\text{Tr } S_n(I - P_m) \leq 4^{-m}$, and hence

$$\int \text{Tr}(S_1 \otimes S_2)(I - P_m) \pi_n(dS_1 dS_2) \leq 4^{-m}. \quad (11.46)$$

Introduce the following subsets in the direct product $\mathfrak{S}(\mathcal{H}_1) \times \mathfrak{S}(\mathcal{H}_2)$:

$$K_m = \{(S_1, S_2) : \text{Tr}(S_1 \otimes S_2)(I - P_m) \leq \delta^{-1} 2^{-m}\}; \quad \mathcal{K}_\delta = \bigcap_{m \geq 1} K_m,$$

where $\delta > 0$. By the same Theorem 11.2, the set \mathcal{K}_δ is compact by construction. Its complement satisfies

$$\pi_n(\overline{\mathcal{K}_\delta}) \leq \sum_m \pi_n(\overline{K_m}) \leq \delta \sum_m 2^m \int \text{Tr}(S_1 \otimes S_2)(I - P_m) \pi_n(dS_1 dS_2) \leq \delta$$

according to (11.46). Thus, the sequence of measures $\{\pi_n\}$ satisfies the criterion of weak relative compactness (see [164]), which completes the proof of the theorem. \square

Definition 11.28. Channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$ is called *entanglement-breaking* if for an arbitrary Hilbert space \mathcal{H}_R and an arbitrary state $S \in \mathfrak{S}(\mathcal{H}_A \otimes \mathcal{H}_R)$, the state $(\Phi \otimes \text{Id}_R)[S] \in \mathfrak{S}(\mathcal{H}_B \otimes \mathcal{H}_R)$, where Id_R is the identity map in $\mathfrak{T}(\mathcal{H}_R)$, is separable.

To describe the structure of such channels, we will need a generalization of the notion of an observable (Definition 2.9) to the case of an arbitrary set of outcomes.

Definition 11.29. Let \mathcal{X} be a measurable space with a σ -algebra of measurable subsets \mathcal{B} . An *observable with values in \mathcal{X}* is a *probability operator-valued measure* (POVM) on \mathcal{X} , i.e. a family $M = \{M(B), B \in \mathcal{B}\}$ of Hermitian operators in \mathcal{H} satisfying the conditions

- i. $M(B) \geq 0; \quad B \in \mathcal{B}$
- ii. $M(\mathcal{X}) = I$

- iii. for any countable decomposition $B = \cup B_j$, ($B_i \cap B_j = \emptyset, i \neq j$) one has $M(B) = \sum_j M(B_j)$ in the sense of weak operator convergence

A probability operator-valued measure E is called *orthogonal* (or a *spectral measure*) if

$$E(B_1 \cap B_2) = E(B_1)E(B_2); \quad B_1, B_2 \in \mathcal{B}.$$

In this case, all operators $E(B)$ are projectors, and the projectors that correspond to disjoint subsets are orthogonal. The corresponding observable is called *sharp*.

The probability distribution of the observable M in the state S is given by the probability measure

$$\mu_S^M(B) = \text{Tr } SM(B), \quad B \in \mathcal{B}. \quad (11.47)$$

Notice that the linear extension of the affine map $S \rightarrow \mu_S^M$ can be regarded as a generalization of the notion of a quantum-classical (q-c) channel (see Section 6.4).

Exercise 11.30. Prove that formula (11.47) defines a probability measure on \mathcal{B} .

Theorem 11.31. *The channel Φ is entanglement-breaking if and only if there is a complete, separable metric space \mathcal{X} , a Borel $\mathfrak{S}(\mathcal{H}_B)$ -valued function $S_B(x)$, and an observable M in \mathcal{H}_A with values in \mathcal{X} given by POVM $M(dx)$ such that*

$$\Phi[S] = \int_{\mathcal{X}} S_B(x) \mu_S^M(dx). \quad (11.48)$$

Relation (11.48) is a continual version of representation (6.28), and to some extent the statement can be regarded as an infinite-dimensional generalization of Proposition 6.22, saying that entanglement-breaking channel is a concatenation of a q-c channel (measurement of observable M) and a c-q channel (preparation of states $S_B(x)$ depending on the measurement outcome). In the infinite-dimensional case, the difference is that an observable with a non-discrete set of outcomes cannot define a quantum channel in the sense of Definition 11.16. Indeed, in the finite-dimensional case, q-c channels are characterized by the property that all the output states commute. Let Φ be an infinite-dimensional channel with this property. This can then always be represented in the form (11.48), with a *discrete* set \mathcal{X} . There exists a basis $\{|k\rangle\}$ such that all commuting density operators $\Phi[S]$ are diagonal. In this case,

$$\Phi[S] = \sum_k |k\rangle\langle k| \Phi[S] |k\rangle\langle k| = \sum_k |k\rangle\langle k| \text{Tr } SM_k,$$

where $M_k = \Phi^*[|k\rangle\langle k|]$ are the components of an observable with a discrete set of outcomes $\mathcal{X} = \{k\}$.

Sketch of the proof of Theorem 11.31. Let the channel have the form (11.48). Consider a state $\omega \in \mathfrak{S}(\mathcal{H}_A \otimes \mathcal{H}_R)$. Then

$$(\Phi \otimes \text{Id}_R)(\omega) = \int_{\mathfrak{X}} S_B(x) \otimes m_\omega(dx), \quad (11.49)$$

where

$$m_\omega(B) = \text{Tr}_A \omega(M(B) \otimes I_R), \quad B \subseteq \mathfrak{X}.$$

Any matrix element of the operator-valued measure m_ω (in a particular basis) is a complex-valued measure on \mathfrak{X} , absolutely continuous with respect to the probability measure $\mu_\omega(B) = \text{Tr } m_\omega(B)$, $B \subseteq \mathfrak{X}$. The Radon–Nikodým theorem implies the representation

$$m_\omega(B) = \int_B \sigma_\omega(x) \mu_\omega(dx),$$

where $\sigma_\omega(x)$ is a function on \mathfrak{X} that takes values in $\mathfrak{S}(\mathcal{H}_R)$. By using this representation, we can rewrite (11.49) as

$$(\Phi \otimes \text{Id}_R)(\omega) = \int_{\mathfrak{X}} S'(x) \otimes \sigma_\omega(x) \mu_\omega(dx), \quad (11.50)$$

which, by Proposition 11.27, is a separable state.

Conversely, let Φ be an entanglement-breaking channel. Fix a nondegenerate state σ_A in $\mathfrak{S}(\mathcal{H}_A)$ and let $\{|e_j\rangle; j = 1, \dots\}$ be the basis of the eigenvectors of σ_A with corresponding (positive) eigenvalues $\{\lambda_j\}$. Consider the unit vector

$$|\Omega\rangle = \sum_{j=1}^{+\infty} \sqrt{\lambda_j} |e_j\rangle \otimes |e_j\rangle$$

in the space $\mathcal{H}_A \otimes \mathcal{H}_A$. Then $|\Omega\rangle\langle\Omega|$ is a purification of the state σ_A . Since Φ is entanglement-breaking, the state

$$\sigma_{AB} = (\text{Id}_A \otimes \Phi)[|\Omega\rangle\langle\Omega|] \quad (11.51)$$

in $\mathfrak{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is separable. By (11.43), there exists a probability measure π on $\mathfrak{S}(\mathcal{H}_A) \times \mathfrak{S}(\mathcal{H}_B)$ such that

$$\sigma_{AB} = \int_{\mathfrak{S}(\mathcal{H}_A)} \int_{\mathfrak{S}(\mathcal{H}_B)} S_A \otimes S_B \pi(dS_A dS_B). \quad (11.52)$$

This implies

$$\begin{aligned}
 \sigma_A &= \text{Tr}_B(\text{Id}_A \otimes \Phi)(|\Omega\rangle\langle\Omega|) \\
 &= \int_{\mathfrak{S}(\mathcal{H}_A)} \int_{\mathfrak{S}(\mathcal{H}_B)} S_A \pi(dS_A dS_B) \\
 &= \int_{\mathfrak{S}(\mathcal{H}_A)} \int_{\mathfrak{S}(\mathcal{H}_B)} \bar{S}_A \pi(dS_A dS_B), \tag{11.53}
 \end{aligned}$$

where the bar denotes complex conjugation in the basis $\{|e_i\rangle\}$. By this equality, for an arbitrary Borel set $B \subseteq \mathfrak{S}(\mathcal{H}_B)$, the operator

$$M(B) = \sigma_A^{-1/2} \left[\int_{\mathfrak{S}(\mathcal{H}_A)} \int_B \bar{S}_A \pi(dS_A dS_B) \right] \sigma_A^{-1/2} \tag{11.54}$$

can be correctly defined as a bounded positive operator on \mathcal{H}_A such that $M(B) \leq M(\mathcal{X}) = I_A$. It is easy to check that M is an observable with values in $\mathcal{X} = \mathfrak{S}(\mathcal{H}_B)$.

Now, consider the entanglement-breaking channel

$$\hat{\Phi}(S) = \int_{\mathfrak{S}(\mathcal{H}_B)} S_B \mu_S^M(dS_B),$$

and let us show that $\Phi(S) = \hat{\Phi}(S)$. For this it is sufficient to prove that

$$\hat{\Phi}(|e_i\rangle\langle e_j|) = \Phi(|e_i\rangle\langle e_j|)$$

for all i, j . However,

$$\begin{aligned}
 \hat{\Phi}(|e_i\rangle\langle e_j|) &= \int_{\mathfrak{S}(\mathcal{H}_B)} S_B \langle e_j | M(dS_B) | e_i \rangle \\
 &= \lambda_i^{-1/2} \lambda_j^{-1/2} \int_{\mathfrak{S}(\mathcal{H}_A)} \int_{\mathfrak{S}(\mathcal{H}_B)} \langle e_i | S_A | e_j \rangle S_B \pi(dS_A dS_B) \\
 &= \lambda_i^{-1/2} \lambda_j^{-1/2} \text{Tr}_A(|e_j\rangle\langle e_i| \otimes I_B) \sigma_{AB} = \Phi(|e_i\rangle\langle e_j|),
 \end{aligned}$$

where in the last equality relation (11.51) was used. □

As was shown in Section 8.3.3, in finite dimensions, entanglement-breaking channels form a large class, in which the additivity conjecture for the classical capacity holds. This fact can be generalized to the infinite-dimensional case.

Proposition 11.32. *Let there be two channels Φ_1, Φ_2 with the corresponding input constraints F_1, F_2 , satisfying the conditions of Corollary 11.24, and let Φ_1 be an entanglement-breaking channel. Then all additivity properties (8.36), (8.33), and (8.39) hold. In particular, the classical capacity of an entanglement-breaking channel Φ is equal to*

$$C(\Phi, F, E) = C_\chi(\Phi, F, E).$$

The proof of this statement is obtained by generalizing the argument to the finite-dimensional case (Proposition 8.19), with the replacement of sums with integrals and ensembles with generalized ensembles, taking into account that the conditions of Corollary 11.24 guarantee finiteness of all involved entropies.

11.8 Notes and references

1. The first attempt at a mathematically rigorous treatment of the basic notions of Quantum Mechanics in a separable Hilbert space was made in the classical treatise of von Neumann [212]. In particular, the difference between Hermitian (symmetric) and self-adjoint operators, and the key role of self-adjointness plays in the spectral decomposition were stressed, in contrast with the preceding works by physicists. The other important circle of ideas is related to the notions of trace and trace-class operators. A modern treatment of these concepts, in connection with applications in quantum theory, is given in the books [171], [48], and [107].

Lemma 11.1 is due to Dell'Antonio [49], see also Appendix to the book of Davies [48]. The compactness criterion for subsets of quantum states is a modification of a result of Sarymsakov [175], called by him the “noncommutative Prokhorov’s Theorem”. The latter provides a criterion for weak compactness of families of probability measures on a metric space, see [164]. It is used in Section 11.4.

2. It is well known that the topological properties of entropy in the infinite-dimensional case differ sharply from those in finite dimensions. In the last case, the entropy is a bounded, continuous function on $\mathfrak{S}(\mathcal{H})$, while in infinite dimensions it is discontinuous everywhere and infinite “almost everywhere” in the sense that the set of states with finite entropy is a first-category subset of $\mathfrak{S}(\mathcal{H})$. See the survey of Wehrl [215], where a number of useful properties of the entropy, including Theorem 11.6 and Lemma 11.8, are discussed. For recent results concerning the topological properties of the entropy in infinite dimensions, see Shirokov [183].

3. The importance of considering input constraints for quantum channels was clear from the beginnings of quantum communication. For a detailed physical discussion of the problem of the determination of the capacity of an optical channel with constrained energy of the input signal, see the survey [38]. The present study of the classical capacity of a constrained c-q channel is based on the works [98] and [102],

taking advantage of the random encoding (11.21), which is used in a similar classical problem.

4. Channels with a continuous alphabet were considered in the works of Holevo [102] and Holevo and Shirokov [110], where the notion of a generalized ensemble was systematically applied. Proofs of the used facts from the theory of probability measures on metric spaces, such as Exercise 11.12 and Lemma 11.13, can be found in the monograph of Parthasarathy [164].

5. For the proof of the boundedness of the map Φ in Exercise 11.17, see the book [48], Lemma 2.2.1. Our consideration of the χ -capacity of constrained quantum channels is based on the works [102] and [110]. A detailed investigation of the properties of the χ -capacity and other entropic characteristics of infinite-dimensional channels, without assuming finiteness of the output entropy, was made by Shirokov [185], [184].

6. This section is based on the results of the work [102].

7. The concept of a quantum observable as POVM is presented in detail in the books [107], [99]. Theorem 11.31 is proved in the paper of Holevo, Shirokov, and Werner [111]. Concerning the proof of Proposition 11.32, see [106].

Chapter 12

Gaussian systems

The fundamental physical information carrier is the electromagnetic field as exemplified by light or radio waves. Mathematically, the radiation field is known to be equivalent to an ensemble of oscillators. In quantum optics, one considers the quantized field and hence quantum oscillators. This is a typical “continuous variable” bosonic quantum system, whose basic observables (oscillator amplitudes) satisfy the canonical commutation relations (CCR). Many of the current experimental realizations of quantum information processing are carried out in such systems.

There is a class of particularly important states of bosonic systems which naturally correspond to classical multidimensional Gaussian distributions. From a physical viewpoint, this class comprises thermal equilibrium states of light as well as coherent and squeezed states produced by lasers and some “nonlinear” quantum optics devices. Mathematically, they are completely characterized by the mean and the covariance matrix, in close parallel with the classical case. Restricting ourselves to a finite number of oscillator modes, which is the usual approximation in quantum optics, makes it possible to handle them with techniques from finite-dimensional linear algebra.

Our central problems will be to unravel the structure of quantum Gaussian channels and to find their capacities. Both problems appear rather involved mathematically and are up to now only partially solved (especially as concerns the capacity problem). Along with the partial results, we provide formulations of the basic conjectures, which, we hope, will stimulate attempts at their solutions.

Unavoidably, a number of analytical complications related to infinite dimensionality and the unboundedness of operators arise in connection with bosonic systems and Gaussian states. In our treatment of CCR, we focus on the aspects essential to applications, while a detailed presentation of the related mathematical tools can be found in the literature. For the convenience of nonspecialist readers, some mathematical and physical preliminaries, as well as motivation, are given in the introductory section.

12.1 Preliminary material

12.1.1 Spectral decomposition and Stone’s Theorem

Let \mathcal{H} be a separable Hilbert space and let $\{E(B); B \in \mathcal{B}(\mathbb{R})\}$ be a spectral measure on \mathbb{R} , where $\mathcal{B}(\mathbb{R})$ is the σ -algebra of Borel subsets. According to Definition 11.29,

the spectral measure E defines a sharp real observable. As explained below, such observables are in one-to-one correspondence with self-adjoint operators in \mathcal{H} .

For an arbitrary unit vector $\psi \in \mathcal{H}$, the relation

$$\mu_\psi(B) = \langle \psi | E(B) | \psi \rangle$$

defines the Borel probability measure on \mathbb{R} . If $f(x)$ is a Borel measurable function on \mathbb{R} , the integral

$$X_f = \int_{-\infty}^{\infty} f(x) E(dx) \quad (12.1)$$

converges strongly on the dense domain

$$\mathcal{D}(X_f) = \left\{ \psi : \int_{-\infty}^{\infty} |f(x)|^2 \mu_\psi(dx) < \infty \right\}$$

and thus, uniquely defines a linear operator on the domain $\mathcal{D}(X_f)$ (strong convergence of operators $\{X_n\}$ to the operator X on the domain \mathcal{D} means that $\lim_{n \rightarrow \infty} \|X_n \psi - X \psi\| = 0$ for $\psi \in \mathcal{D}$. For the case where $\mathcal{D} = \mathcal{H}$ it is just the *strong operator convergence*.)

Indeed, if $f(x) = \sum_{i=1}^{\infty} f_i 1_{B_i}(x)$ is a countably-valued function (here $\{B_i\}$ is a measurable decomposition of \mathbb{R}), then

$$X_f = \sum_{i=1}^{\infty} f_i E(B_i),$$

and due to the orthogonality of the vectors $E(B_i)|\psi\rangle$,

$$\|X_f \psi\|^2 = \sum_{i=1}^{\infty} |f_i|^2 \langle \psi | E(B_i) | \psi \rangle = \sum_{i=1}^{\infty} |f_i|^2 \mu_\psi(B_i), \quad (12.2)$$

so that $\mathcal{D}(X_f)$ is just the set of all vectors $|\psi\rangle$ for which this series converges. This definition can be extended to arbitrary Borel function $f(x)$ by approximating it uniformly with countably-valued functions and using (12.2). In particular, for the function $f(x) = x$, we obtain the operator

$$X = \int_{-\infty}^{\infty} x E(dx) \quad (12.3)$$

with the domain

$$\mathcal{D}(X) = \left\{ \psi : \int_{-\infty}^{\infty} |x|^2 \mu_\psi(dx) < \infty \right\},$$

for which

$$\langle \psi | X | \psi \rangle = \int_{-\infty}^{\infty} x \mu_\psi(dx); \quad \|X \psi\|^2 = \int_{-\infty}^{\infty} |x|^2 \mu_\psi(dx), \quad \psi \in \mathcal{D}(X). \quad (12.4)$$

The operators obtained in this way are self-adjoint in the following sense. For any densely defined operator X , there is a unique adjoint X^* satisfying

$$\langle X^* \varphi | \psi \rangle = \langle \varphi | X \psi \rangle, \quad \psi \in \mathcal{D}(X), \varphi \in \mathcal{D}(X^*), \quad (12.5)$$

where $\mathcal{D}(X^*)$ is the (dense) subspace of all vectors φ such that the right-hand side is a bounded linear functional of ψ . Operator X is called *Hermitian* (symmetric) if

$$\langle X \varphi | \psi \rangle = \langle \varphi | X \psi \rangle, \quad (12.6)$$

for all $\varphi, \psi \in \mathcal{D}(X)$ and *self-adjoint* if $X = X^*$ in the sense that $\mathcal{D}(X^*) = \mathcal{D}(X)$ and (12.6) holds. Often, one has to deal with *essentially self-adjoint* operators defined on some non-maximal domain which, however, extend uniquely to self-adjoint operators.

The *Spectral Theorem* (see e.g. [171], Section VIII.3) says that for any self-adjoint operator X there exists a unique spectral measure E on \mathbb{R} for which the relations (12.3), (12.4) hold. Thus, self-adjoint operators are those which have the spectral decomposition (12.3) over the real line. Such operators are naturally associated with (sharp) real observables in Quantum Mechanics. In this case, the measure μ_ψ is the probability distribution of the observable X in the pure state $|\psi\rangle\langle\psi|$, and the formulas (12.4) provide the first and the second moments of this distribution.

In a certain sense, on which we will not comment here, the relation $X_f = f(X)$ holds. By taking $f(x) = \exp(itx)$ in (12.1), we obtain the operators

$$V_t = \int_{-\infty}^{\infty} \exp(itx) E(dx) = \exp itX.$$

Then the family $\{V_t; t \in \mathbb{R}\}$ is a strongly continuous group of unitary operators with the infinitesimal generator X :

$$\lim_{t \rightarrow 0} (it)^{-1} (V_t - I) |\psi\rangle = X |\psi\rangle; \quad \psi \in \mathcal{D}(X). \quad (12.7)$$

Conversely, any strongly continuous group of unitary operators $\{V_t; t \in \mathbb{R}\}$ has the form $V_t = \exp itX$, where X is a uniquely defined self-adjoint operator. These statements are the content of *Stone's Theorem*.

Exercise 12.1. 1. Let $\mathcal{H} = L^2(\mathbb{R})$ be the Hilbert space of square integrable functions $\psi(\xi)$, $\xi \in \mathbb{R}$, and $E = \{E(B); B \in \mathcal{B}(\mathbb{R})\}$ be defined by the expression

$$(E(B)\psi)(\xi) = 1_B(\xi) \psi(\xi); \quad \psi \in \mathcal{H}.$$

Prove that E is the spectral measure of the operator q of multiplication by ξ , while V_t is the operator of multiplication by $\exp it\xi$ and, in general, X_f is the operator of multiplication by $f(\xi)$.

2. Consider the unitary operator in \mathcal{H} defined by the Fourier transform

$$(\tilde{F}\psi)(\lambda) = \frac{1}{\sqrt{2\pi}} \int \exp(-i\lambda\xi) \psi(\xi) d\xi.$$

Then $\tilde{E}(B) = \tilde{F}^* E(B) \tilde{F}$ is a spectral measure on \mathbb{R} , corresponding to the self-adjoint operator $p \equiv \tilde{q} = \frac{1}{i} \frac{d}{d\xi}$, which generates the group of unitary operators

$$(U_t \psi)(\xi) = \psi(\xi + t).$$

Self-adjoint operators X_j ; $j = 1, \dots, s$ commute if their spectral measures commute. The following is a multidimensional version of Stone's Theorem.

Theorem 12.2. *Let V_x ; $x = [x_1, \dots, x_s] \in \mathbb{R}^s$ be a strongly continuous group of unitary operators in a separable Hilbert space \mathcal{H} . Then there exists a family X_j ; $j = 1, \dots, s$ of commuting self-adjoint operators in \mathcal{H} such that*

$$V_x = \exp\left(i \sum_{j=1}^s x_j X_j\right). \quad (12.8)$$

Conversely, for any family X_j ; $j = 1, \dots, s$ of commuting self-adjoint operators in \mathcal{H} relation (12.8) defines a strongly continuous unitary group in \mathcal{H} .

12.1.2 Operators associated with the Heisenberg commutation relation

In the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$ we consider the operators

$$(q\psi)(\xi) = \xi \psi(\xi); \quad (p\psi)(\xi) = \frac{\hbar}{i} \frac{d}{d\xi} \psi(\xi),$$

defined on a common dense domain $\mathcal{D} \subseteq \mathcal{H}$. For example, \mathcal{D} can be the subspace $\mathcal{S}(\mathbb{R})$ of infinitely differentiable, rapidly decreasing functions, with all derivatives tending to zero quicker than any degree of $|\xi|$ when $|\xi| \rightarrow \infty$. These operators are essentially self-adjoint. Hence, they represent (sharp) real observables (see Section 12.1.1). Here \hbar is a positive constant equal to Planck's constant in physical applications, where it relates different systems of units (we will later choose the system in which $\hbar = 1$).

On this domain, q and p satisfy the *Heisenberg commutation relation*

$$[q, p] = i\hbar I. \quad (12.9)$$

For a unit vector ψ from the common domain of q, p , consider a pure state $|\psi\rangle\langle\psi|$ and introduce notation

$$x = \langle\psi|q|\psi\rangle, \quad y = \langle\psi|p|\psi\rangle, \\ D_\psi(q) = \|(q - x)\psi\|^2, \quad D_\psi(p) = \|(p - y)\psi\|^2$$

for the mean values and the variances of observables q, p . Then for any real ω

$$\begin{aligned}\omega^2 D_\psi(q) - \omega\hbar + D_\psi(p) &= \omega^2 D_\psi(q) + i\omega \langle \psi | [q, p] | \psi \rangle + D_\psi(p) \\ &= \| [\omega(q - x) + i(p - y)] \psi \|^2 \geq 0,\end{aligned}$$

whence

$$D_\psi(q)D_\psi(p) \geq \frac{\hbar^2}{4}, \quad (12.10)$$

with the equality attained if and only if there exists a positive ω such that

$$[\omega(q - x) + i(p - y)] \psi = 0.$$

This amounts to the differential equation

$$\left[\omega(\xi - x) + \left(\hbar \frac{d}{d\xi} - iy \right) \right] \psi(\xi) = 0, \quad (12.11)$$

the normalized solution of which, up to a constant factor of modulus one, is

$$\psi(\xi) = \sqrt[4]{\frac{\omega}{\pi\hbar}} \exp \left[\frac{iy}{\hbar} \left(\xi - \frac{x}{2} \right) - \frac{\omega(\xi - x)^2}{2\hbar} \right]. \quad (12.12)$$

Inequality (12.10) is the *Heisenberg uncertainty relation* and (12.12) are the *minimal uncertainty* state vectors, with $D_\psi(q) = \frac{\hbar}{2\omega}$, $D_\psi(p) = \frac{\omega\hbar}{2}$.

By introducing the complex combinations

$$a = \frac{1}{\sqrt{2\hbar\omega}} (\omega q + ip); \quad \zeta = \frac{1}{\sqrt{2\hbar\omega}} (\omega x + iy)$$

and denoting the corresponding state vector (12.12) by $|\zeta\rangle$, we can rewrite the defining equation (12.11) as

$$a|\zeta\rangle = \zeta|\zeta\rangle; \quad \zeta \in \mathbb{C}. \quad (12.13)$$

In particular,

$$a|0\rangle = 0, \quad (12.14)$$

where $|0\rangle$ is the vector described by the function

$$\psi(\xi) = \sqrt[4]{\frac{\omega}{\pi\hbar}} \exp \left[-\frac{\omega\xi^2}{2\hbar} \right].$$

The operator $a = \frac{1}{\sqrt{2\hbar\omega}} (\omega q + ip)$ has the adjoint $a^\dagger = \frac{1}{\sqrt{2\hbar\omega}} (\omega q - ip)$. The commutation relation (12.9) can be rewritten as

$$[a, a^\dagger] = I. \quad (12.15)$$

Consider the operator

$$\mathcal{N} = a^\dagger a = aa^\dagger - I, \quad (12.16)$$

which is essentially self-adjoint. By (12.14) it satisfies

$$\mathcal{N}|0\rangle = 0.$$

Exercise 12.3. Successively applying relation (12.16), show that the vectors

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle \quad (12.17)$$

form an orthonormal system of eigenvectors of the operator \mathcal{N} :

$$\mathcal{N}|n\rangle = n|n\rangle; \quad n = 0, 1, \dots$$

The corresponding functions in $L^2(\mathbb{R})$, up to normalizing factors, are

$$\left[\omega\xi - \hbar \frac{d}{d\xi} \right]^n \exp \left[-\frac{\omega\xi^2}{2\hbar} \right] \simeq H_n \left(\sqrt{\frac{\omega}{\hbar}} \xi \right) \exp \left[-\frac{\omega\xi^2}{2\hbar} \right],$$

where $H_n(\cdot)$; $n = 0, 1, \dots$ are the Hermite polynomials, which form a complete orthogonal system in $L^2(\mathbb{R})$. Hence, $\{|n\rangle; n = 0, 1, \dots\}$ is an orthonormal basis in \mathcal{H} ,

$$\sum_{n=0}^{\infty} |n\rangle\langle n| = I,$$

and \mathcal{N} has the spectral decomposition

$$\mathcal{N} = \sum_{n=0}^{\infty} n|n\rangle\langle n|. \quad (12.18)$$

Exercise 12.4. Operator \mathcal{N} is self-adjoint of the type \mathfrak{F} (Definition 11.3) with the domain

$$\mathcal{D}(\mathcal{N}) = \left\{ \psi : \sum_{n=0}^{\infty} n^2 |\langle n|\psi\rangle|^2 < \infty \right\}$$

and (12.18) holds in the sense of strong convergence on $\mathcal{D}(\mathcal{N})$.

Further, the relations

$$a|n\rangle = \sqrt{n}|n-1\rangle; \quad a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$$

hold. By introducing the group of unitary operators $\{e^{i\mathcal{N}\omega t}; t \in \mathbb{R}\}$, one has

$$e^{i\mathcal{N}\omega t} a e^{-i\mathcal{N}\omega t} = a e^{-i\omega t}; \quad e^{i\mathcal{N}\omega t} a^\dagger e^{-i\mathcal{N}\omega t} = a^\dagger e^{i\omega t}. \quad (12.19)$$

Coming back to the operators q, p , one sees that

$$\hbar\omega\left(\mathcal{N} + \frac{I}{2}\right) = \frac{1}{2}(\omega^2 q^2 + p^2) \equiv \hbar H \quad (12.20)$$

is the *energy operator* for the quantum harmonic oscillator with frequency ω . Then \mathcal{N} is the *number (of quanta) operator*, $|n\rangle$ are the number state vectors, $|0\rangle$ is the *vacuum* state vector, a (resp. a^\dagger) is the *annihilation* (resp. *creation*) operator. In quantum optics the operators a, a^\dagger describe a mode of the electromagnetic field corresponding to a definite frequency (and a definite polarization).

12.1.3 Classical signal plus quantum noise

The states $|\zeta\rangle\langle\zeta|; \zeta \in \mathbb{C}$, are called *coherent*. From (12.17), (12.13), we obtain

$$\langle n|\zeta\rangle = \frac{1}{\sqrt{n!}}\langle 0|a^n\zeta\rangle = \frac{\zeta^n}{\sqrt{n!}}\langle 0|\zeta\rangle = \frac{\zeta^n}{\sqrt{n!}}\exp\left(-\frac{|\zeta|^2}{2}\right),$$

where we also used the formula

$$\langle\zeta_1|\zeta_2\rangle = \exp\left[-\frac{1}{2}(|\zeta_1|^2 + |\zeta_2|^2 - 2\bar{\zeta}_1\zeta_2)\right]. \quad (12.21)$$

Exercise 12.5. Prove (12.21) by using the real parametrization (12.12) of the vectors $|\zeta\rangle$.

Exercise 12.6. The system of vectors $\{|\zeta\rangle; \zeta \in \mathbb{C}\}$ is *overcomplete* in the sense

$$\frac{1}{\pi} \int |\zeta\rangle\langle\zeta| d^2\zeta = I,$$

where $d^2\zeta = \frac{1}{2\hbar} dx dy$. Hint: the matrix elements of the integral in the basis $\{|n\rangle\}$ have the form

$$\frac{1}{\pi} \int \frac{\zeta^n}{\sqrt{n!}} \frac{\bar{\zeta}^m}{\sqrt{m!}} \exp(-|\zeta|^2) d^2\zeta = \delta_{nm}.$$

Let $p(\zeta)$ be a probability density on \mathbb{C} . Then

$$S = \int |\zeta\rangle\langle\zeta| p(\zeta) d^2\zeta \quad (12.22)$$

is a density operator of a state called *classical* in quantum optics. In particular, taking the complex Gaussian density with zero mean and variance N , we obtain the density operator

$$S_0 = \frac{1}{\pi N} \int |\zeta\rangle\langle\zeta| \exp\left(-\frac{|\zeta|^2}{N}\right) d^2\zeta \quad (12.23)$$

with the matrix elements

$$\langle n|S_0|m\rangle = \frac{1}{\pi N} \int \frac{\zeta^n}{\sqrt{n!}} \frac{\bar{\zeta}^m}{\sqrt{m!}} \exp\left(-\frac{(N+1)|\zeta|^2}{N}\right) d^2\zeta = \delta_{nm} \frac{1}{N+1} \left(\frac{N}{N+1}\right)^n.$$

Therefore, the operator S_0 has the spectral decomposition

$$S_0 = \frac{1}{N+1} \sum_{n=0}^{\infty} \left(\frac{N}{N+1}\right)^n |n\rangle\langle n|. \quad (12.24)$$

This relation also makes sense for $N = 0$ if S_0 is defined by the Gaussian distribution that is degenerate at the point $\zeta = 0$.

Taking into account the spectral decomposition (12.18) and relation (12.20), this can be written as

$$S_0 = \frac{\exp[-\theta H]}{\text{Tr} \exp[-\theta H]}, \quad (12.25)$$

which is the Gibbs equilibrium (*thermal*) state of the quantum harmonic oscillator at the inverse temperature $\theta/\hbar = \frac{1}{\hbar\omega} \ln \frac{N+1}{N}$. The parameter $N = \text{Tr} S_0 a^\dagger a$ is interpreted as the mean number of the oscillator energy quanta. The von Neumann entropy of the state (12.24)

$$H(S_0) = \frac{1}{N+1} \sum_{n=0}^{\infty} \left(\frac{N}{N+1}\right)^n [(n+1) \log(N+1) - n \log N] = g(N), \quad (12.26)$$

where we introduced the function

$$g(x) = (x+1) \log(x+1) - x \log x, \quad x > 0; \\ g(0) = 0.$$

Another important density operator is obtained from the shifted probability density

$$S_\mu = \frac{1}{\pi N} \int |\zeta\rangle\langle\zeta| \exp\left(-\frac{|\zeta - \mu|^2}{N}\right) d^2\zeta, \quad (12.27)$$

where $\mu = \frac{1}{\sqrt{2\hbar\omega}} (\omega m_q + i m_p)$. Introducing the unitary operators

$$D(\zeta)\psi(\xi) = \exp\left[\frac{iy}{\hbar} \left(\xi - \frac{x}{2}\right)\right] \psi(\xi - x) \quad (12.28)$$

we see that

$$\begin{aligned} |\zeta\rangle &= D(\zeta)|0\rangle; \\ D(\zeta_2)|\zeta_1\rangle &= \exp(i\Im\bar{\zeta}_1\zeta_2)|\zeta_1 + \zeta_2\rangle. \end{aligned}$$

That is, the action of this operator is to make a displacement in the mean values of q, p . Physically, such a displacement is realized by the action of an external source, such as an idealized laser producing a coherent state $|\zeta\rangle\langle\zeta|$ from the vacuum state $|0\rangle\langle 0|$. The quasiclassical state (12.22) is a chaotic mixture of coherent states, while

$$S_\mu = D(\mu)S_0D(\mu)^* \quad (12.29)$$

represents the transformation of the equilibrium state under the action of an external source characterized by the complex amplitude μ . Notice that the state S_μ is pure if and only if $N = 0$, in which case it coincides with the coherent state $|\mu\rangle\langle\mu|$. In applications to information theory, μ is the classical signal, to be transmitted via the noisy quantum mode q, p . The state S_μ thus is a model for a classical signal plus quantum Gaussian noise (see the next section).

Exercise 12.7. Prove that the displacement operators satisfy the commutation relation

$$D(\zeta_2)D(\zeta_1) = \exp(i\Im\bar{\zeta}_1\zeta_2)D(\zeta_1 + \zeta_2), \quad (12.30)$$

which implies

$$D(\mu)^*D(\zeta)D(\mu) = \exp(2i\Im\bar{\mu}\zeta)D(\zeta). \quad (12.31)$$

In what follows, we need the noncommutative analog of the characteristic function for the state S_μ :

$$\mathrm{Tr} S_\mu D(\zeta) = \exp\left[2i\Im\bar{\mu}\zeta - \left(N + \frac{1}{2}\right)|\zeta|^2\right]. \quad (12.32)$$

Proof of formula (12.32). By (12.29), (12.31) we have

$$\mathrm{Tr} S_\mu D(\zeta) = \exp(2i\Im\bar{\mu}\zeta) \mathrm{Tr} S_0 D(\zeta)$$

and it is sufficient to prove (12.32) for $\mu = 0$. Now,

$$\begin{aligned} \mathrm{Tr} S_0 D(\zeta) &= \int \langle\zeta_1|D(\zeta)|\zeta_1\rangle \frac{1}{\pi N} \exp\left(-\frac{|\zeta_1|^2}{N}\right) d^2\zeta_1 \\ &= \int \langle 0|D(\zeta_1)^*D(\zeta)D(\zeta_1)|0\rangle \frac{1}{\pi N} \exp\left(-\frac{|\zeta_1|^2}{N}\right) d^2\zeta_1 \\ &= \int \exp(i\Im\bar{\zeta}_1\zeta) \langle 0|D(\zeta)|0\rangle \frac{1}{\pi N} \exp\left(-\frac{|\zeta_1|^2}{N}\right) d^2\zeta_1 \end{aligned} \quad (12.33)$$

By using (12.21), we obtain

$$\langle 0|D(\zeta)|0\rangle = \langle 0|\zeta\rangle = \exp\left(-\frac{1}{2}|\zeta|^2\right).$$

Substituting this expression into (12.33) and using the expression for the characteristic function of the Gaussian probability distribution, we obtain

$$\text{Tr } S_0 D(\zeta) = \exp\left(-\left(N + \frac{1}{2}\right)|\zeta|^2\right).$$

□

12.1.4 The classical-quantum Gaussian channel

The mapping $\mu \rightarrow S_\mu$ can be considered as a classical-quantum (c-q) channel in the sense of Section 11.4, realizing the transmission of the classical signal $\mu \in \mathbb{C}$ with the additive quantum Gaussian noise of power N . For the first two moments, we have

$$\text{Tr } S_\mu a = \mu, \quad \text{Tr } S_\mu a^\dagger a = N + |\mu|^2. \quad (12.34)$$

To compute its classical capacity, we can use the general approach of Section 8.2.

Assuming that the words $w = (\mu_1, \dots, \mu_n)$ are transmitted by independent uses of this channel (memoryless channel), we impose the additive energy constraint on the signal μ of type (11.16),

$$|\mu_1|^2 + \dots + |\mu_n|^2 \leq nE, \quad (12.35)$$

which corresponds to the constraint function $f(\mu) = |\mu|^2$. In this case, the conditions of Theorem 11.15 are satisfied with the choice $F = a^\dagger a \equiv \mathcal{N}$, and the classical capacity of the channel $\mu \rightarrow S_\mu$ with the energy input constraint (12.35) is equal to

$$C = \max_{\pi \in \mathcal{P}_E} \left[H(\bar{S}_\pi) - \int H(S_\mu) \pi(d^2\mu) \right], \quad (12.36)$$

where \mathcal{P}_E is the set of input distributions $\pi(d^2\mu)$ satisfying

$$\int |\mu|^2 \pi(d^2\mu) \leq E, \quad (12.37)$$

and

$$\bar{S}_\pi = \int S_\mu \pi(d^2\mu).$$

Note that, by (12.29), the states S_μ are unitarily equivalent and hence have the same entropy $H(S_\mu) = H(S_0) = g(N)$, see (12.26). Therefore, for any input distribution $\pi(d^2\mu)$, the χ -quantity (11.24) is equal to

$$\chi(\pi) = H(\bar{S}_\pi) - H(S_0),$$

and the problem reduces to maximization of the entropy $H(\bar{S}_\pi)$ under the constraint (12.37). Due to (12.34), relation (12.37) implies

$$\text{Tr } \bar{S}_\pi a^\dagger a \leq N + E. \quad (12.38)$$

This is a restriction onto the second moments of the state \bar{S}_π . According to the maximal entropy principle (see Lemma 12.25 below), the maximal value of the entropy

$$H(\bar{S}_\pi) = g(N + E), \quad (12.39)$$

is attained by the Gaussian density operator

$$\bar{S}_\pi = \frac{1}{N + E + 1} \sum_{n=0}^{\infty} \left(\frac{N + E}{N + E + 1} \right)^n |n\rangle\langle n|, \quad (12.40)$$

which fulfils the condition (12.38) with equality sign. It corresponds to the optimal distribution

$$\pi(d^2\mu) = \frac{1}{\pi E} \exp\left(-\frac{|\mu|^2}{E}\right) d^2\mu. \quad (12.41)$$

Finally, the capacity of the memoryless c-q Gaussian channel is given by the expression

$$\begin{aligned} C &= C_\chi = g(N + E) - g(N) \\ &= \log\left(1 + \frac{E}{N + 1}\right) + (N + E) \log\left(1 + \frac{1}{N + E}\right) - N \log\left(1 + \frac{1}{N}\right), \end{aligned} \quad (12.42)$$

which behaves as $\log\left(1 + \frac{E}{N+1}\right)$, asymptotically in the limit $N \rightarrow \infty$, $E/N \rightarrow \text{const}$. Thus, relation (12.42) can be regarded as the quantum generalization of Shannon's formula (4.51), with Gaussian white noise of power N . The factor 1/2 is absent in (12.42), because one oscillator mode corresponds to the two independent identically distributed real amplitudes m_q, m_p .

In what follows, we will develop a general theory of quantum Gaussian channels and their capacities.

12.2 Canonical commutation relations

12.2.1 Weyl-Segal CCR

In quantum mechanics, the canonical commutation relations (CCR) arise in the quantization of a mechanical system with a finite number of degrees of freedom, or a classical field represented as an infinite collection of oscillators. In quantum optics, one usually deals only with a finite number of relevant oscillator frequencies, thus again reducing to the case of mechanical system with a finite number s of degrees of freedom.

Consider the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^s)$ of complex, square-integrable functions of the real variables ξ_j ; $j = 1, \dots, s$, where \mathbb{R}^s is the coordinate space of the underlying classical system. In the space \mathcal{H} , we consider the two groups of unitary operators

$$V_x \psi(\xi) = \exp(i \xi^\top x) \psi(\xi); \quad U_y \psi(\xi) = \psi(\xi + y), \quad (12.43)$$

where $\xi, x, y \in \mathbb{R}^s$ are understood as column vectors and $^\top$ denotes transposition. These groups satisfy the Weyl CCR

$$U_y V_x = \exp(i y^\top x) V_x U_y. \quad (12.44)$$

Notice the analogy with the discrete groups arising from (6.51). The groups U_y, V_x describe the change of a quantum state under the displacements in position (respectively, momentum) space, and the Weyl CCR is an expression for the kinematics of a nonrelativistic quantum mechanical system, which in fact can be derived from the Galilei covariance, see e.g. [107].

Computing the generators of the groups V_x, U_y by a multidimensional analog of formula (12.7), we obtain a particular instance of Stone's Theorem

$$V_x = \exp\left(i \sum_{j=1}^s x_j q_j\right), \quad U_y = \exp\left(i \sum_{j=1}^s y_j p_j\right),$$

where $x = [x_1 \dots x_s]^\top$, $y = [y_1 \dots y_s]^\top$ and the self-adjoint operators

$$q_j = \xi_j, \quad p_j = \frac{1}{i} \frac{\partial}{\partial \xi_j}$$

are the *canonical observables* which satisfy the Heisenberg CCR

$$[q_j, p_k] = i \delta_{jk} I, \quad [q_j, q_k] = 0, \quad [p_j, p_k] = 0. \quad (12.45)$$

Here and in what follows, we choose units in which $\hbar = 1$. The operators q_j, p_j are unbounded. Hence, relations (12.45) should be considered only on a common dense domain, such as the subspace $\mathcal{S}(\mathbb{R})$.

To make use of the apparent symmetry between x, y , we introduce the $2s$ -vector $z = [x_1, y_1, \dots, x_s, y_s]^\top$ and the unitary Weyl operators

$$W(z) = \exp\left(\frac{i}{2} y^\top x\right) V_x U_y. \quad (12.46)$$

From (12.46) and (12.44) it follows that the operators $W(z)$ satisfy the *Weyl–Segal CCR*

$$W(z)W(z') = \exp\left[-\frac{i}{2} \Delta(z, z')\right] W(z + z'), \quad (12.47)$$

where

$$\Delta(z, z') = \sum_{j=1}^s (x_j y'_j - x'_j y_j) = z^\top \Delta z' \quad (12.48)$$

is the canonical *symplectic form*. Here, we denote by the same letter the matrix of the form

$$\Delta = \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & & & \\ & & \ddots & & \\ & & & 0 & 1 \\ & & & -1 & 0 \end{bmatrix} \equiv \text{diag} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \quad (12.49)$$

From (12.47) and the unitarity of the operators $W(z)$ it follows that $W(-z) = W(z)^*$. Relation (12.47) implies

$$W(z')^* W(z) W(z') = \exp[-i \Delta(z, z')] W(z). \quad (12.50)$$

The similarity between the commutation relations (12.30) and (12.47) is not accidental. In fact, the operators $W(-\Delta^{-1}z)$ are the multimode generalization of the unitary displacement operators.

Exercise 12.8. Show that in the case of one degree of freedom

$$D(\mu) = W(m_p/\hbar, -m_q/\hbar), \quad (12.51)$$

where $\mu = \frac{1}{\sqrt{2\hbar\omega}}(\omega m_q + i m_p)$, so that

$$D(\mu)^* W(z) D(\mu) = \exp i(m_q x + m_p y) W(z). \quad (12.52)$$

Since $\Delta(z, z) \equiv 0$, relation (12.47) implies that for any fixed z the one-parameter family $\{W(tz); t \in \mathbf{R}\}$ is a unitary group. The generator of this group, computed according to the formula (12.7), is the self-adjoint operator

$$Rz = \sum_{j=1}^s (x_j q_j + y_j p_j),$$

where

$$R = [q_1 \ p_1 \ \dots \ q_s \ p_s]$$

is the row vector of the canonical observables. According to Stone's Theorem,

$$W(z) = \exp i Rz. \quad (12.53)$$

Exercise 12.9. Use (12.50) to prove the relation

$$W(-\Delta^{-1}z)^* R W(-\Delta^{-1}z) = R + z^\top I. \quad (12.54)$$

Exercise 12.10. Show that the Heisenberg commutation relations (12.45) can be written in the form

$$[Rz, Rz'] = i \Delta(z, z') I. \quad (12.55)$$

12.2.2 The symplectic space

The space $Z = \mathbb{R}^{2s}$, equipped with the nondegenerate skew-symmetric form $\Delta(z, z')$, is a *symplectic vector space*. It represents the phase space of the classical system, the quantum version of which is described by the family of unitary operators $W(z)$ in the Hilbert space \mathcal{H} . This quantization is essentially unique. Any irreducible family of unitary operators $W(z)$ in a Hilbert space satisfying Weyl–Segal CCR is unitarily equivalent to the representation in $L^2(\mathbb{R}^s)$ described above, which is called the Schrödinger representation (Stone–von Neumann’s Uniqueness Theorem, see e.g. [171], [107]). In what follows, $W(z); z \in Z$ can be any irreducible representation of the CCR.

A basis $\{e_j, h_j; j = 1, \dots, s\}$ in Z is called *symplectic* if

$$\Delta(e_j, h_k) = \delta_{jk}, \Delta(e_j, e_k) = \Delta(h_j, h_k) = 0; \quad j, k = 1, \dots, s. \quad (12.56)$$

In any such basis, the symplectic form $\Delta(z, z')$ has the standard expression (12.48). The transition matrix T from the initial symplectic basis in Z to the new symplectic basis is a matrix of *symplectic transformation* in (Z, Δ) , which is characterized by the property

$$\Delta(Tz, Tz') = \Delta(z, z'); \quad z, z' \in Z.$$

The operator J in (Z, Δ) is called an *operator of complex structure* if

$$J^2 = -1, \quad (12.57)$$

where 1 is the identity operator in Z , and it is Δ -positive in the sense that the bilinear form

$$j(z, z') = \Delta(z, Jz') \quad (12.58)$$

is an inner product in Z . Note that Δ -positivity is equivalent to the conditions

$$\Delta J = -J^\top \Delta, \quad \Delta J \geq 0. \quad (12.59)$$

Exercise 12.11. Prove the following statement: operator J defines the structure of a complex unitary space in Z (of dimensionality s), in which $iz = Jz$ and the inner product is

$$j(z, z') + i\Delta(z, z') = \Delta(z, Jz') + i\Delta(z, z').$$

In what follows, we will consider various bilinear forms α, \dots in the space $Z = \mathbb{R}^{2s}$, and the matrices of such forms will be denoted by the same letters, e.g. $\alpha(z, z') = z^\top \alpha z'$, etc.

Lemma 12.12. Let $\alpha(z, z') = z^\top \alpha z'$ be an inner product in the symplectic space (Z, Δ) . Then there is a symplectic basis $\{e_j, h_j; j = 1, \dots, s\}$ in Z such that the form α is diagonal, with the matrix

$$\tilde{\alpha} = \text{diag} \begin{bmatrix} \alpha_j & 0 \\ 0 & \alpha_j \end{bmatrix}, \quad (12.60)$$

where $\alpha_j > 0$.

Proof. Consider the operator $A = \Delta^{-1}\alpha$, satisfying

$$\alpha(z, z') = \Delta(z, Az').$$

The operator A is skew-symmetric in the Euclidean space $(Z, \alpha) : A^* = -A$. According to a theorem from linear algebra, there is an orthogonal basis $\{e_j, h_j\}$ in (Z, α) and positive numbers $\{\alpha_j\}$, such that

$$Ae_j = \alpha_j h_j; \quad Ah_j = -\alpha_j e_j.$$

Choosing the normalization $\alpha(e_j, e_j) = \alpha(h_j, h_j) = \alpha_j$ produces the symplectic basis in (Z, Δ) , with the required properties. \square

For arbitrary inner product α in Z , there is at least one operator of complex structure J , commuting with the operator $A = \Delta^{-1}\alpha$, namely, the orthogonal operator J from the polar decomposition

$$A = |A| J = J |A| \quad (12.61)$$

in the Euclidean space (Z, α) . Applying Lemma 12.12, we obtain that there is a symplectic basis $\{e_j, h_j; j = 1, \dots, s\}$ in which the form α is diagonal with the matrix (12.60), while J has the matrix

$$\tilde{J} = \text{diag} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad (12.62)$$

so that

$$Je_j = h_j, \quad Jh_j = -e_j.$$

Denoting by T the transition matrix from the initial symplectic basis in Z to the new symplectic basis, i.e. the matrix with the columns $\{e_j, h_j; j = 1, \dots, s\}$, we have

$$\Delta = T^\top \Delta T; \quad \tilde{\alpha} = T^\top \alpha T; \quad \tilde{J} = T^{-1} J T.$$

The canonical observables in the new basis are given by the relations

$$\tilde{q}_j = R e_j, \quad \tilde{p}_j = R h_j; \quad j = 1, \dots, s,$$

so that $RT = \tilde{R}$, where $\tilde{R} = [\tilde{q}_1 \ \tilde{p}_1 \ \dots \ \tilde{q}_s \ \tilde{p}_s]$. The complex structure is most simply expressed in terms of the *creation - annihilation operators*

$$\tilde{a}_j^\dagger = \frac{1}{\sqrt{2}} (\tilde{q}_j - i \tilde{p}_j), \quad \tilde{a}_j = \frac{1}{\sqrt{2}} (\tilde{q}_j + i \tilde{p}_j).$$

Namely, the action of the operator J on the canonical observables $R \rightarrow RJ$ is expressed as the multiplication

$$\tilde{a}_j^\dagger \rightarrow i \tilde{a}_j^\dagger, \quad \tilde{a}_j \rightarrow -i \tilde{a}_j; \quad j = 1, \dots, s. \quad (12.63)$$

This follows from the fact that $[\tilde{q}_j \ \tilde{p}_j] \tilde{J} = [\tilde{p}_j \ -\tilde{q}_j]$.

12.2.3 Dynamics, quadratic operators and gauge transformations

For any symplectic transformation T in Z , there is a unitary operator U_T in \mathcal{H} such that

$$U_T^* W(z) U_T = W(Tz). \quad (12.64)$$

This follows from Stone-von Neumann's Uniqueness Theorem because the operators $W(Tz); z \in Z$, again form an (irreducible) representation of the CCR (12.47) in \mathcal{H} . In view of relation (12.53), this is equivalently expressed in terms of canonical observables

$$U_T^* R U_T = R T. \quad (12.65)$$

Let us consider one-parameter semigroups of symplectic transformations in Z

$$T_t = e^{tD}; \quad t \in \mathbb{R},$$

that describe linear dynamics in the classical phase space. We assume that the generator D is Δ -positive, i.e. $\Delta(z, Dz')$ is an inner product in Z . In matrix notations,

$$\Delta D = -D^\top \Delta \geq 0. \quad (12.66)$$

The quantization of the linear dynamics T_t is given by the following Theorem.

Theorem 12.13. Consider the operator $H = R\epsilon R^\top$ in \mathcal{H} , which is quadratic in the canonical variables R , where $\epsilon = -\frac{1}{2}D\Delta^{-1}$. Then H is a positive self-adjoint operator in \mathcal{H} generating the unitary group $\{e^{itH}\}$ such that

$$W(T_t z) = e^{itH} W(z) e^{-itH}. \quad (12.67)$$

Proof (sketch). In terms of generators of the Weyl unitaries, the relation (12.67) that we wish to establish is equivalent to

$$R e^{tD} = e^{itR\epsilon R^\top} R e^{-itR\epsilon R^\top}, \quad (12.68)$$

with $D = -2\epsilon\Delta$.

Consider the bilinear form

$$\frac{1}{2}\Delta(z, Dz') = -z^\top \Delta \epsilon \Delta z'$$

which, by our assumption about the operator D , is an inner product. According to Lemma 12.12, there is a symplectic basis $\{e_j, h_j\}$ such that the matrix of this form is diagonal

$$-\Delta \tilde{\epsilon} \Delta = \text{diag} \begin{bmatrix} \omega_j/2 & 0 \\ 0 & \omega_j/2 \end{bmatrix},$$

where $\omega_j > 0$. Hence,

$$\tilde{\epsilon} = \text{diag} \begin{bmatrix} \omega_j/2 & 0 \\ 0 & \omega_j/2 \end{bmatrix}.$$

By introducing the new canonical observables $\tilde{q}_j = R e_j$, $\tilde{p}_j = R h_j$; $j = 1, \dots, s$, we have

$$H = \sum_{j=1}^s \frac{\omega_j}{2} (\tilde{q}_j^2 + \tilde{p}_j^2), \quad (12.69)$$

implying that H is a positive self-adjoint operator as a sum of operators of the type (12.20) referring to different modes. In terms of the creation-annihilation operators,

$$H = \sum_{j=1}^s \omega_j (\tilde{a}_j^\dagger \tilde{a}_j + I/2). \quad (12.70)$$

In the new basis $\{e_j, h_j\}$, the matrix of operator $D = -2\epsilon\Delta$ has the form

$$\tilde{D} = \text{diag} \begin{bmatrix} 0 & -\omega_j \\ \omega_j & 0 \end{bmatrix} = \tilde{J} \begin{bmatrix} \omega_j & 0 \\ 0 & \omega_j \end{bmatrix},$$

where the last expression is just the polar decomposition of \tilde{D} , so that the complex structure $J = T \tilde{J} T^{-1}$ arises from the similar decomposition

$$D = J|D| = |D|J. \quad (12.71)$$

Consider formula (12.68) that we wish to establish in the new basis. Taking into account the second relation in (12.63), which shows that the right action of J on R is equivalent to multiplication of the annihilation operators by $-i$, this formula reduces to equations of the type (12.19):

$$\tilde{a}_j e^{-i\omega_j t} = e^{i\omega_j t} \tilde{a}_j^\dagger \tilde{a}_j \tilde{a}_j e^{-i\omega_j t} \tilde{a}_j^\dagger \tilde{a}_j; \quad j = 1, \dots, s, \quad (12.72)$$

describing the dynamics of quantum harmonic oscillators with frequencies ω_j . \square

The components of the quantum system described by the operators $\tilde{a}_j, \tilde{a}_j^\dagger$ or \tilde{q}_j, \tilde{p}_j , corresponding to the frequencies ω_j , are called *normal modes*, and the approach based on the representation of the energy matrix ϵ in the diagonal form is just the normal mode decomposition.

With every complex structure, we can associate the cyclic one-parameter group $\{e^{\varphi J}; \varphi \in [0, 2\pi]\}$ of symplectic transformations, which we call the *gauge group*. According to the above Theorem, the gauge group in Z induces the unitary group of *gauge transformations* in \mathcal{H} by the formula

$$W(e^{\varphi J} z) = e^{-i\varphi G} W(z) e^{i\varphi G}, \quad (12.73)$$

where $G = \frac{1}{2} R J \Delta^{-1} R^T$ is a positive self-adjoint operator in \mathcal{H} . In terms of generators, (12.73) reduces to the canonical transformation

$$R e^{\varphi J} = e^{-i\varphi G} R e^{i\varphi G}, \quad (12.74)$$

which is a particular case of (12.68).

Applying the previous argument to the case where $D = J$ and using (12.57), we obtain that there exists a symplectic basis $\{e_j, h_j; j = 1, \dots, s\}$ such that $J e_j = h_j, J h_j = -e_j$, in which

$$G = \sum_{j=1}^s \frac{1}{2} (\tilde{q}_j^2 + \tilde{p}_j^2) = \mathcal{N} + \frac{s}{2} I, \quad (12.75)$$

where

$$\mathcal{N} = \sum_{j=1}^s \tilde{a}_j^\dagger \tilde{a}_j$$

is the *total number operator*.

An operator X in \mathcal{H} is called *gauge-invariant* if

$$e^{-i\varphi G} X e^{i\varphi G} = X$$

for all $\varphi \in [0, 2\pi]$. By using (12.74) and (12.59), we find that a quadratic operator $X = R \epsilon R^T$, where ϵ is a symmetric positive matrix, is gauge-invariant if

$$[J, \epsilon \Delta] = 0,$$

i.e. J is the operator of complex structure from the polar decomposition (12.71) of $D = -2\epsilon\Delta$. During the proof of Theorem 12.13 we established that such a complex structure exists for every energy matrix ϵ .

Exercise 12.14. Let ω_j ; $j = 1, \dots, s$ be positive numbers, and

$$H = \sum_{j=1}^s \frac{1}{2} (\omega_j^2 q_j^2 + p_j^2)$$

be the Hamiltonian of the ensemble of oscillators with the frequencies ω_j , then $\tilde{q}_j = \sqrt{\omega_j} q_j$, $\tilde{p}_j = \sqrt{\omega_j}^{-1} p_j$, so that

$$\tilde{a}_j^\dagger = \frac{1}{\sqrt{2\omega_j}} (\omega_j q_j - i p_j), \quad \tilde{a}_j = \frac{1}{\sqrt{2\omega_j}} (\omega_j q_j + i p_j)$$

and

$$H = \sum_{j=1}^s \frac{\omega_j}{2} (\tilde{q}_j^2 + \tilde{p}_j^2) = \sum_{j=1}^s \omega_j (\tilde{a}_j^\dagger \tilde{a}_j + I/2).$$

The corresponding operator of complex structure is

$$J = \text{diag} \begin{bmatrix} 0 & -\omega_j \\ \omega_j^{-1} & 0 \end{bmatrix},$$

while the gauge operator G in \mathcal{H} is given by (12.75).

12.3 Gaussian states

12.3.1 Characteristic function

The *characteristic function* of a quantum state S is defined as

$$\phi_S(z) = \text{Tr } S W(z); \quad z \in Z.$$

This is a kind of noncommutative Fourier transform, uniquely defining the operator S . The corresponding inversion formula is

$$S = \frac{1}{(2\pi)^s} \int \phi_S(z) W(-z) d^{2s}z,$$

where $d^{2s}z = dx_1 \dots dy_s$ is the element of symplectic volume in Z . Similar relations hold for an arbitrary trace class operator S , see [107].

Positivity of the operator S implies (and is in fact equivalent to) the nonnegative definiteness of $n \times n$ -matrices:

$$\left[\phi(z_r - z_s) \exp \left(\frac{i}{2} \Delta(z_r, z_s) \right) \right]_{r,s=1,\dots,n} \geq 0 \quad (12.76)$$

for all $n = 1, 2, \dots$ and arbitrary collections $\{z_1, \dots, z_n\} \in Z$. To see the necessity of this condition, notice that the matrix element is nothing but $\text{Tr } W(z_s)^* S W(z_r)$ so that

$$\sum_{r,s} c_r \bar{c}_s \phi(z_r - z_s) \exp\left(\frac{i}{2} \Delta(z_r, z_s)\right) = \text{Tr } S A A^*,$$

where $A = \sum_r c_r W(z_r)$.

A density operator S has *finite second moments* if $\text{Tr } S(q_j^2 + p_j^2) < \infty$ for all j , where the trace is defined as in (11.6). In this case, one can define the *mean vector* $\text{Tr } S R = m$ and the *covariance matrix* $B_S(R) = \alpha$ (see Section 2.3.3). By CCR (12.55) the commutation matrix $C_S(R) = -\Delta$, so that

$$\alpha + \frac{i}{2} \Delta = \text{Tr } (R - m)^\top S (R - m). \quad (12.77)$$

This relation (and its transpose) implies the inequality

$$\alpha \geq \pm \frac{i}{2} \Delta, \quad (12.78)$$

which is nothing but the uncertainty relation (2.20) for the canonical observables R . In Theorem 12.17 below we show that condition (12.78) is not only necessary, but also sufficient for α to be the covariance matrix of a quantum state. We denote by $\Sigma(m, \alpha)$ the set of all states with the mean vector m and the covariance matrix α .

As in probability theory, the components of m, α (as well as higher moments) can be expressed via the derivatives of the characteristic function.

Exercise 12.15. Assuming the existence of the corresponding moments, show that

$$\text{Tr } S(Rz)^n = i^{-n} \frac{d^n}{dt^n} \phi(tz) \Big|_{t=0}.$$

12.3.2 Definition and properties of Gaussian states

The characteristic function (12.32) of the one-mode density operator (12.27), expressed in real variables, using (12.51), is

$$\text{Tr } S W(z) = \exp\left[i(m_q x + m_p y) - \frac{\alpha}{2}(x^2 + y^2)\right], \quad (12.79)$$

where $\alpha = N + \frac{1}{2} \geq \frac{1}{2}$. We call such a state S *elementary Gaussian state*. Let us now elaborate the general definition of a multimode Gaussian state.

The state S is called *Gaussian*, if its characteristic function $\phi(z) = \text{Tr } S W(z)$ has the form

$$\phi(z) = \exp\left(i m(z) - \frac{1}{2} \alpha(z, z)\right), \quad (12.80)$$

where $m(z) = mz$ is a linear form and $\alpha(z, z) = z^T \alpha z$ is a bilinear form on Z , Δ . Here, m is a row $2s$ -vector and α is a real symmetric $(2s) \times (2s)$ -matrix.

Exercise 12.16. Considering the derivatives of (12.80) at $z = 0$, show that m is indeed the mean vector, and α is the covariance matrix determined from (12.77).

Usually, (m, α) are called the parameters of the Gaussian state. In the case $m = 0$ we call the Gaussian state *centered*. If S_m is the Gaussian state with parameters (m, α) , then

$$S_m = W(-\Delta^{-1}m^T)S_0W(-\Delta^{-1}m^T)^*, \quad (12.81)$$

as follows from (12.54). By this unitary equivalence many questions reduce to the consideration of centered states.

Theorem 12.17. For relation (12.80) to define a quantum state, it is necessary and sufficient that the matrix α satisfies condition (12.78). Under this condition, formula (12.80) defines the unique Gaussian state in $\Sigma(m, \alpha)$.

Proof. It was already mentioned that the necessity follows from relation (12.77). An alternative proof follows from the next lemma, which will also be needed in the sequel.

Lemma 12.18. Let α be an inner product, Δ a skew-symmetric form in Z , such that for all $n = 1, 2, \dots$ and arbitrary collection $\{z_1, \dots, z_n\} \in Z$,

$$\left[\exp \left(\alpha(z_r, z_s) + \frac{i}{2} \Delta(z_r, z_s) \right) \right]_{r,s=1,\dots,n} \geq 0 \quad (12.82)$$

In this case, the matrices of forms α, Δ satisfy condition (12.78).

Proof. Let $t > 0$. By writing the condition of nonnegative definiteness for the collection $\{0, \sqrt{t}z_1, \dots, \sqrt{t}z_n\}$ with the variables $\{c_0, c_1, \dots, c_n\} \in \mathbb{C}$ such that $c_0 = -\sum_{j=1}^n c_j$, we have

$$\sum_{r,s=1}^n \bar{c}_r c_s \left(\exp t \left[\alpha(z_r, z_s) + \frac{i}{2} \Delta(z_r, z_s) \right] - 1 \right) \geq 0.$$

Dividing by t and letting $t \rightarrow 0$ we obtain that the matrices with elements $\alpha(z_r, z_s) + \frac{i}{2} \Delta(z_r, z_s)$ are nonnegative definite for an arbitrary collection z_1, \dots, z_n , which is equivalent to condition (12.78). This proves the lemma. \square

Substituting the Gaussian expression (12.80) into (12.76), one can see that

$$\phi(z_r - z_s) \exp \left[\frac{i}{2} \Delta(z_r, z_s) \right] = \phi(z_r) \overline{\phi(z_s)} \exp \{ \alpha(z_r, z_s) + \frac{i}{2} \Delta(z_r, z_s) \}.$$

Thus, the condition of Lemma 12.18 is satisfied, implying inequality (12.78).

To prove sufficiency, assume that the matrix α satisfies inequality (12.78). Lemma 12.12 then implies that there exists a symplectic transformation T such that

$$\tilde{\alpha} = T^\top \alpha T = \text{diag} \begin{bmatrix} \alpha_j & 0 \\ 0 & \alpha_j \end{bmatrix}, \quad (12.83)$$

where $\alpha_j > 0$, while inequality (12.78) is easily seen to be equivalent to

$$\alpha_j \geq \frac{1}{2}, \quad j = 1, \dots, s. \quad (12.84)$$

Relation (12.83) implies

$$\begin{aligned} \phi(Tz) &= \exp(i\tilde{m}z - \frac{1}{2}z^\top \tilde{\alpha}z) \\ &= \exp \sum_{j=1}^s [i(\tilde{m}_{jq}x_j + \tilde{m}_{jp}y_j) - \frac{1}{2}\alpha_j(x_j^2 + y_j^2)], \end{aligned}$$

where $\tilde{m} = mT$.

Denoting by $S^{(j)}$ the corresponding elementary one-mode Gaussian states (12.79) and putting

$$S = \bigotimes_{j=1}^s S^{(j)}, \quad (12.85)$$

we obtain

$$\phi(Tz) = \text{Tr } S \exp i \tilde{R}z = \text{Tr } S W(Tz),$$

where $\tilde{R} = RT$ are the new canonical observables. It follows that the state S has the characteristic function $\phi(z)$. \square

In quantum optics, the representation of the Gaussian state in the form (12.85) is usually related to the normal mode decomposition.

Let J be an operator of complex structure in Z and $\{e^{i\varphi G}\}$ be the corresponding gauge group in \mathcal{H} . From (12.73) it follows that the Gaussian density operator S is gauge-invariant, $e^{-i\varphi G} S e^{i\varphi G} = S$, $\varphi \in [0, 2\pi]$, if and only if its characteristic function satisfies the condition $\phi(e^{\varphi J}z) = \phi(z)$, which is equivalent to $m = 0$ and $J^\top \alpha + \alpha J = 0$. By using (12.59), the last equality can be written as

$$[J, A] = 0, \quad (12.86)$$

where $A = \Delta^{-1}\alpha$, which means that J is the complex structure from the polar decomposition (12.61), i.e. $J = T\tilde{J}T^{-1}$, where \tilde{J} is defined by relation (12.62).

Notice that the elementary Gaussian state (12.79) is pure if and only if $N = 0$, i.e. $\alpha = 1/2$. Using the decomposition in (12.85) this implies several equivalent conditions for purity of a general Gaussian state.

Exercise 12.19. A Gaussian state S is pure if and only if one of the following equivalent conditions holds:

- i. $\alpha_j = 1/2; \quad j = 1, \dots, s$
- ii. $|\det(2A)| = 1$
- iii. the operator $A = \Delta^{-1}\alpha$ satisfies

$$A^2 = -\frac{1}{4}I \quad (12.87)$$

- iv. $\alpha = \frac{1}{2}\Delta J$, where J is an operator of complex structure
- v. α is a minimal solution of the inequality (12.78)

Hint: It is sufficient to check conditions ii.–iv. in the symplectic basis from Lemma 12.12.

Under these conditions, the set $\Sigma(m, \alpha)$ consists of one pure Gaussian state.

Let S be a centered Gaussian state with the covariance matrix α , let J be an operator of complex structure from the polar decomposition (12.61), and let S_0 be the pure centered Gaussian state with the covariance matrix $\frac{1}{2}\Delta J$. Thus, both these states are gauge-invariant under the unitary group $\{\exp(i\varphi G)\}$, corresponding to this complex structure.

Exercise 12.20. By using the normal mode decomposition (12.85), prove the following multimode generalization of formula (12.23)

$$S = \int W(z) S_0 W(z)^* P(d^{2s}z), \quad (12.88)$$

where P is the Gaussian probability distribution with the characteristic function

$$\int e^{i\Delta(w,z)} P(d^{2s}z) = \exp \left[-w^\top \left(\alpha - \frac{1}{2}\Delta J \right) w \right].$$

This distribution is invariant under the complex structure J in Z .

The gauge-invariant pure state S_0 plays the role of the *vacuum state* while $W(-\Delta^{-1}z)S_0W(-\Delta^{-1}z)^*$ are the *coherent states* with respect to the complex structure J . Note that a state that is “squeezed” relative to a given complex structure can always be made “coherent” by using the complex structure associated with its covariance matrix. In physics, the distinguished complex structure is one with which the oscillator Hamiltonian has the canonical form (12.69) and coherent/squeezed states are defined relative to that complex structure.

12.3.3 The density operator of Gaussian state

The spectral decomposition of an arbitrary Gaussian state (12.85) can be obtained by tensor multiplication of the spectral decompositions of the one-mode states $S^{(j)}$ from the normal mode decomposition. We will prove a theorem that generalizes relation (12.25) to arbitrary Gaussian states without a pure component.

Lemma 12.21. *The Gaussian density operator S is nondegenerate if and only if*

$$\det \left(\alpha + \frac{i}{2} \Delta \right) \neq 0.$$

Proof. The Gaussian density operator S is nondegenerate if and only if it has no pure component, i.e. $\alpha_j > \frac{1}{2}$; $j = 1, \dots, s$. We will prove the lemma by establishing the identities

$$\det \left(\alpha + \frac{i}{2} \Delta \right) = \left[\det \left(-A^2 - \frac{1}{4} I \right) \right]^{\frac{1}{2}} = \prod_{j=1}^s \left(\alpha_j^2 - \frac{1}{4} \right), \quad (12.89)$$

where $A = \Delta^{-1} \alpha$. Choosing the symplectic basis as in Lemma 12.12, we have $A = T(\Delta^{-1} \tilde{\alpha}) T^{-1}$, where

$$\Delta^{-1} \tilde{\alpha} = \text{diag} \begin{bmatrix} 0 & -\alpha_j \\ \alpha_j & 0 \end{bmatrix}. \quad (12.90)$$

Hence, the numbers $\pm i\alpha_j$; $j = 1, \dots, s$ are the eigenvalues of the operator A . It follows that the symmetric operator $-A^2 - \frac{1}{4} I$ in the Euclidean space (Z, α) is positive, since its eigenvalues are $\alpha_j^2 - 1/4 \geq 0$ (notice that, by (12.87), this operator is equal to 0 if and only if the state S is pure.). Each of these eigenvalues appears twice. Hence, the second identity in (12.89) follows.

To prove the first identity, note that

$$-(\Delta^{-1} \alpha)^2 - \frac{1}{4} I = -\Delta^{-1} \left(\alpha + \frac{i}{2} \Delta \right) \Delta^{-1} \left(\alpha - \frac{i}{2} \Delta \right).$$

Taking into account that $\det(\pm \Delta^{-1}) = 1$ and $\det(\alpha + \frac{i}{2} \Delta) = \det(\alpha - \frac{i}{2} \Delta) \geq 0$ we obtain the result. \square

Theorem 12.22. *Let the matrix $\alpha + \frac{i}{2} \Delta$ be nondegenerate. In this case, the centered Gaussian density operator with covariance matrix α has the form*

$$S_0 = c \exp \left(-R \epsilon R^T \right), \quad (12.91)$$

where

$$c = \left[\det \left(\alpha + \frac{i}{2} \Delta \right) \right]^{-\frac{1}{2}} = \left[\det (-4 \sin^2 \epsilon \Delta) \right]^{\frac{1}{4}}, \quad (12.92)$$

and ϵ is found from the relation

$$2\Delta^{-1}\alpha = \cot \epsilon \Delta. \quad (12.93)$$

Proof. Rewriting the relation (12.25) (for $\omega = 1$) in the form

$$\left(e^{\theta/2} - e^{-\theta/2}\right) \exp\left(-\frac{\theta}{2}(q^2 + p^2)\right),$$

using the normal modes decomposition (12.83) diagonalizing α and (12.85), we have

$$S_0 = c \exp\left(-\tilde{R}\tilde{\epsilon}\tilde{R}^\top\right),$$

where

$$c = \prod_{j=1}^s \left(e^{\theta_j/2} - e^{-\theta_j/2}\right), \quad \tilde{\epsilon} = \text{diag} \begin{bmatrix} \theta_j/2 & 0 \\ 0 & \theta_j/2 \end{bmatrix}$$

and $\theta_j = \ln \left(\frac{\alpha_j + \frac{1}{2}}{\alpha_j - \frac{1}{2}}\right)$, so that

$$e^{\theta_j/2} - e^{-\theta_j/2} = \left(\alpha_j^2 - \frac{1}{4}\right)^{-\frac{1}{2}}. \quad (12.94)$$

Taking into account (12.89), the first of the expressions (12.92) follows. Returning to the initial canonical observables R , we obtain (12.91), where $\epsilon = T\tilde{\epsilon}T^\top$.

Inverting the formula (12.94) produces

$$\alpha_j = \frac{1}{2} \coth \frac{\theta_j}{2}. \quad (12.95)$$

Further,

$$\tilde{\epsilon}\Delta = \text{diag} \begin{bmatrix} 0 & \theta_j/2 \\ -\theta_j/2 & 0 \end{bmatrix}.$$

Note that $\epsilon\Delta = T\tilde{\epsilon}\Delta T^{-1}$ and $\Delta^{-1}\alpha = T\Delta^{-1}\tilde{\alpha}T^{-1}$ are the matrices of operators, with $\Delta^{-1}\tilde{\alpha}$ given by the relation (12.90). The operators $-\tilde{\epsilon}\Delta$ and $\tilde{\Delta}^{-1}\tilde{\alpha}$ have the same eigenvectors, with eigenvalues $\pm i\theta_j/2$ and $\pm i\alpha_j$. Therefore, via relation $\cot(i\lambda) = -i \coth \lambda$, (12.95) implies (12.93). \square

12.3.4 Entropy of a Gaussian state

To compute the von Neumann entropy of a general Gaussian state, one can use the normal mode decomposition (12.85). By (12.81), the entropy is the same for all Gaussian states with arbitrary mean and covariance matrix α , reducing the problem to the case of a centered state. For a single mode ($s = 1$), the density operator $S^{(j)}$ is unitarily

equivalent to the state (12.24), with entropy equal to $g(N)$. For a general Gaussian state we obtain, by summing over normal modes,

$$H(S) = \sum_{j=1}^s g(N_j), \quad (12.96)$$

where $N_j = \alpha_j - 1/2 \geq 0$.

To write this in coordinate-free form, recall that the operator $A = \Delta^{-1}\alpha$ has eigenvalues $\pm i\alpha_j$. Hence, its matrix is diagonalizable (in the complex domain). For any diagonalizable matrix $M = U \text{diag}(m_j) U^{-1}$, we set $\text{abs}(M) = U \text{diag}(|m_j|) U^{-1}$, similar to other continuous functions on the complex plane. Now, equation (12.96) can be written as

$$H(S) = \frac{1}{2} \text{Sp } g \left(\text{abs}(\Delta^{-1}\alpha) - \frac{I}{2} \right), \quad (12.97)$$

where Sp is used to denote the trace of a matrix, as distinct from the trace Tr of operators in the underlying Hilbert space.

An alternative expression for the entropy of a Gaussian state can be obtained from the representation (12.91). By definition, we have

$$H(S) = -\log c + \text{Sp} \epsilon \alpha,$$

which, by (12.92) and (12.93), is equal to

$$\begin{aligned} & \frac{1}{4} \text{Sp} \log \left[-(\Delta^{-1}\alpha)^2 - \frac{1}{4} I \right] \\ & + \text{Sp}(\Delta^{-1}\alpha) \text{arccot}(2\Delta^{-1}\alpha) = -\frac{1}{4} \text{Sp} \log(-\sin^2 \epsilon \Delta) + \frac{1}{2} \text{Sp}(\epsilon \Delta \cot \epsilon \Delta). \end{aligned} \quad (12.98)$$

Exercise 12.23. Use the relation

$$g\left(x - \frac{1}{2}\right) = \frac{1}{2} \log\left(x^2 - \frac{1}{4}\right) + x \log \frac{x + \frac{1}{2}}{x - \frac{1}{2}},$$

to show that (12.97) and (12.98) are the same.

Example 12.24. In the case of a one-mode Gaussian density operator, we have $\alpha = \begin{bmatrix} \alpha^{qq} & \alpha^{qp} \\ \alpha^{qp} & \alpha^{pp} \end{bmatrix}$ with $\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2 \geq \frac{1}{4}$ (this inequality is equivalent to condition (12.78)). In this case,

$$-(\Delta^{-1}\alpha)^2 = [\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Hence,

$$|\Delta^{-1}\alpha| = \sqrt{\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$H(S) = g \left(\sqrt{\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2} - \frac{1}{2} \right).$$

Geometrically, the quantity $\pi \sqrt{\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2}$ is equal to the area of the deviation ellipsoid

$$z^\top \alpha^{-1} z = 1, \quad z = [x, y]^\top,$$

for the two-dimensional Gaussian distribution with covariance matrix α .

Note that, in accordance with (12.87), the Gaussian state is pure if and only if it has the minimal uncertainty $\alpha^{qq}\alpha^{pp} - (\alpha^{qp})^2 = \frac{1}{4}$. This comprises both coherent states for which

$$\alpha^{qq} = \alpha^{pp} = \frac{1}{2}, \quad \alpha^{qp} = 0$$

and squeezed states for which this condition is violated (as distinct from quantum optics we use here real rather than complex parametrization). The case of many modes can be treated by using the normal mode decomposition.

While pure Gaussian states are the minimal uncertainty states for the canonical observables, a general Gaussian state is characterized by the following property of maximal entropic uncertainty under fixed moments of the canonical observables.

Lemma 12.25. *The Gaussian state has the largest entropy among all states with given mean m and correlation matrix α . Therefore, for any set of states defined by restrictions on the first and second moments, the maximum of the entropy can always be sought among the Gaussian states.*

Proof. Let $S \in \Sigma(m, \alpha)$ and let \tilde{S} be the unique Gaussian state in $\Sigma(m, \alpha)$. Without loss of generality we may assume that \tilde{S} is nondegenerate. The general case can be reduced to this one by separating the pure component in the tensor product decomposition of \tilde{S} . Indeed, by performing the normal modes decomposition, we can assume that $Z = Z_1 \oplus Z_2$ with $\Delta = \Delta_1 \oplus \Delta_2, \alpha = \alpha_1 \oplus \alpha_2$ (see the next section for the direct sum decompositions), where α_1 corresponds to the unique, necessarily pure Gaussian state in Z_1 , while α_2 corresponds to a nondegenerate density operator. We have

$$H(\tilde{S}) - H(S) = H(S; \tilde{S}) + \text{Tr}(S - \tilde{S}) \log \tilde{S},$$

where the last term is equal to zero, because, by (12.91), for a nondegenerate Gaussian \tilde{S} , the operator $\log \tilde{S}$ is a second order polynomial in the canonical variables, while the first and second moments of S, \tilde{S} coincide. Thus, $H(\tilde{S}) - H(S) = H(S; \tilde{S}) \geq 0$. \square

A similar result holds for the conditional quantum entropy [55]. In its formulation and proof, we use the description of composite bosonic systems that we introduced in the next section.

Lemma 12.26. *The Gaussian state has the largest conditional entropy among all states of a composite bosonic system AB with the given mean m and the correlation matrix α .*

Proof. Let $S_{AB} \in \Sigma(m_{AB}, \alpha_{AB})$ and let \tilde{S}_{AB} be the unique Gaussian state in $\Sigma(m_{AB}, \alpha_{AB})$. Then $S_A \in \Sigma(m_A, \alpha_A)$ and \tilde{S}_A is the Gaussian state in $\Sigma(m_A, \alpha_A)$. Again, we may assume that \tilde{S}_{AB} is nondegenerate. Denoting

$$H(A|B) = H(S_{AB}) - H(S_B); \quad H(\tilde{A}|\tilde{B}) = H(\tilde{S}_{AB}) - H(\tilde{S}_B),$$

we have

$$\begin{aligned} H(\tilde{A}|\tilde{B}) - H(A|B) &= H(S_{AB}; \tilde{S}_{AB}) - H(S_A; \tilde{S}_A) \\ &\quad + \text{Tr} (S_{AB} - \tilde{S}_{AB}) \log \tilde{S}_{AB} - \text{Tr} (S_A - \tilde{S}_A) \log \tilde{S}_A \\ &= H(S_{AB}; \tilde{S}_{AB}) - H(S_A; \tilde{S}_A), \end{aligned}$$

because, by (12.91), $\log \tilde{S}_{AB}, \log \tilde{S}_A$ are the second order polynomials in the canonical variables. By monotonicity of the relative entropy, $H(S_{AB}; \tilde{S}_{AB}) - H(S_A; \tilde{S}_A) \geq 0$. \square

12.3.5 Separability and purification

Consider two bosonic systems, described by CCR, with the symplectic spaces $(Z_1, \Delta_1), (Z_2, \Delta_2)$. The symplectic space of the composite system is the direct sum $Z = Z_1 \oplus Z_2$, the elements of which are conveniently represented by the column vectors $z = [z_1, z_2]^\top$, while the symplectic matrix Δ_{12} is block diagonal

$$\Delta_{12} = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}.$$

The Weyl operators of the composite system are defined as $W_{12}(z_1, z_2) = W_1(z_1) \otimes W_2(z_2)$.

Let S_{12} be a Gaussian state of the composite system with the mean m_{12} and the covariance matrix α_{12} . The restriction of the Gaussian state S_{12} to the first factor is determined by the expectations of the Weyl operators $W_1(z_1) \otimes I_2 = W_{12}(z_1, 0)$. Hence, according to (12.80), it has the mean m_1 , which is just the first component of $m_{12} = [m_1 \ m_2]$, and the covariance matrix α_1 , which is the first diagonal block in the block matrix decomposition

$$\alpha_{12} = \begin{bmatrix} \alpha_1 & \beta \\ \beta^\top & \alpha_2 \end{bmatrix}. \quad (12.99)$$

The covariance matrix α_{12} of a state S_{12} of the composite system is block diagonal ($\beta = 0$) if and only if the state is a product $S_{12} = S_1 \otimes S_2$.

Proposition 12.27. [217] *A Gaussian state is separable if and only if there exist real symmetric matrices $\check{\alpha}_j$ satisfying*

$$\check{\alpha}_j \geq \pm \frac{i}{2} \Delta_j; \quad j = 1, 2, \quad (12.100)$$

such that

$$\alpha_{12} \geq \begin{bmatrix} \check{\alpha}_1 & 0 \\ 0 & \check{\alpha}_2 \end{bmatrix}. \quad (12.101)$$

Proof. By using formula (12.81) and the fact that the action of the displacement operators is “local”, and has no effect on (non)separability, we can assume, without loss of generality, that S_{12} is centered. Let S_{12} be separable. In this case, it has a representation (11.43). Since S_{12} has finite second moments, this representation implies that the states $S_1(x)$, $S_2(x)$ have finite second moments for π -almost all x , hence their mean vectors $m_j(x)$ and covariance matrices $\alpha_j(x) \geq \pm \frac{i}{2} \Delta_j$; $j = 1, 2$ are defined for π -almost all x . In view of (12.77), the representation in (11.43) implies that α_{12} has the block decomposition (12.99), with

$$\begin{aligned} \alpha_j &= \int_{\mathcal{X}} \left[\alpha_j(x) + m_j(x)^\top m_j(x) \right] \pi(dx); \quad j = 1, 2, \\ \beta &= \int_{\mathcal{X}} m_1(x)^\top m_2(x) \pi(dx). \end{aligned}$$

Denoting $\check{\alpha}_j = \int_{\mathcal{X}} \alpha_j(x) \pi(dx)$, we have (12.100) and

$$\begin{aligned} \alpha_{12} - \begin{bmatrix} \check{\alpha}_1 & 0 \\ 0 & \check{\alpha}_2 \end{bmatrix} &= \begin{bmatrix} \alpha_1 - \check{\alpha}_1 & \beta \\ \beta^\top & \alpha_2 - \check{\alpha}_2 \end{bmatrix} \\ &= \int_{\mathcal{X}} \begin{bmatrix} m_1(x)^\top m_1(x) & m_1(x)^\top m_2(x) \\ m_2(x)^\top m_1(x) & m_2(x)^\top m_2(x) \end{bmatrix} \pi(dx) \geq 0. \end{aligned}$$

Conversely, if (12.101) holds, let

$$\gamma_{12} = \alpha_{12} - \begin{bmatrix} \check{\alpha}_1 & 0 \\ 0 & \check{\alpha}_2 \end{bmatrix} \geq 0$$

and define π to be the Gaussian probability distribution on $Z = Z_1 \oplus Z_2$ with the characteristic function

$$\int_Z e^{-i \Delta(w, z)} \pi(d^{2s} z) = \exp \left(-\frac{1}{2} z^\top \gamma_{12} z \right).$$

Then comparison of the quantum characteristic functions shows that

$$S_{12} = \int_Z W_1(z_1) S_1 W_1(z_1)^* \otimes W_2(z_2) S_2 W_2(z_2)^* \pi(d^{2s} z), \quad (12.102)$$

where S_j is the centered Gaussian state with covariance matrix $\check{\alpha}_j$. Hence, S_{12} is separable. \square

This proof shows that, in fact, a separable Gaussian state can be represented as a Gaussian mixture of product Gaussian states.

From (12.101) and (12.78) it follows that the covariance matrix of a separable state, in addition to the usual condition $\alpha_{12} \geq \pm \frac{i}{2} \Delta_{12}$, satisfies

$$\alpha_{12} \geq \pm \frac{i}{2} \begin{bmatrix} \Delta_1 & 0 \\ 0 & -\Delta_2 \end{bmatrix}. \quad (12.103)$$

This condition expresses the property of the state S_{12} having a positive partial transpose (PPT) with respect to the second system. In fact, changing the sign of the symplectic form Δ_2 corresponds to taking the transpose of the Weyl operators $W_2(z_2)$, as follows from the CCR (12.47). As shown in [217], the PPT condition is in general weaker than the separability condition (12.101), which means that there exist bound entangled Gaussian states.

When considering purification, it is also sufficient to restrict ourselves to centered Gaussian states. Purification of the one-mode Gaussian state can be performed by using the spectral decomposition (12.24) and the general purification recipe of Theorem 3.11. Then for the many-mode state one can use the decomposition (12.85). This leads to rather lengthly computations [90], and therefore we will only give the final result, and check that it satisfies all necessary requirements.

Now, let S_1 be a Gaussian state, with the covariance matrix α_1 on the symplectic space Z_1 of the form $\Delta_1 = \Delta$. It appears convenient to take $Z_2 = \bar{Z}_1$ with $\Delta_2 = -\Delta$, so that

$$\Delta_{12} = \begin{bmatrix} \Delta & 0 \\ 0 & -\Delta \end{bmatrix}. \quad (12.104)$$

Theorem 12.28. *Consider the symplectic space $Z = Z_1 \oplus \bar{Z}_1$ and the Gaussian state S_{12} with the covariance matrix*

$$\alpha_{12} = \begin{bmatrix} \alpha_1 & \alpha_1^{1/2} B^{-1} \sqrt{-B^2 - I/4} \alpha_1^{1/2} \\ -\alpha_1^{1/2} B^{-1} \sqrt{-B^2 - I/4} \alpha_1^{1/2} & \alpha_1 \end{bmatrix}, \quad (12.105)$$

where $B = \alpha_1^{1/2} \Delta^{-1} \alpha_1^{1/2}$. Then S_{12} is a purification of S_1 .

Proof. We have seen in the proof of Lemma 12.21 that the operator $A = \Delta^{-1} \alpha_1$ is skew-symmetric in the Euclidean space (Z, α) . Moreover, $-A^2 - I/4 \geq 0$ with

equality if and only if the Gaussian state is pure. Consider the block operator

$$A_{12} = \begin{bmatrix} A & \sqrt{-A^2 - I/4} \\ \sqrt{-A^2 - I/4} & -A \end{bmatrix} \quad (12.106)$$

in $Z = Z_1 \oplus \bar{Z}_1$, where the square root is the operator square root in the space (Z, α) . It is easy to check that $-A_{12}^2 - I_{12}/4 = 0$ and, hence, the form $\alpha_{12} = \Delta_{12} A_{12}$ corresponds to the covariance matrix of a pure Gaussian state. Moreover, its restriction to Z_1 coincides with α_1 . Formula (12.105) then gives the matrix expression for α_{12} . Notice that the matrix (12.105) is symmetric non-negative definite, which follows from the fact that $-B^2 - I/4 \geq 0$, while B is skew-symmetric. \square

In the case of the elementary Gaussian state (12.79) in one mode, with

$$\alpha_1 = \begin{bmatrix} N + 1/2 & 0 \\ 0 & N + 1/2 \end{bmatrix}$$

formulas (12.104), (12.105) amount to

$$\Delta_{12} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$\alpha_{12} = \begin{bmatrix} N + 1/2 & 0 & 0 & \sqrt{N^2 + N} \\ 0 & N + 1/2 & -\sqrt{N^2 + N} & 0 \\ 0 & -\sqrt{N^2 + N} & N + 1/2 & 0 \\ \sqrt{N^2 + N} & 0 & 0 & N + 1/2 \end{bmatrix}$$

so that

$$\Delta_{12}^{-1} \alpha_{12} = \begin{bmatrix} 0 & -(N + 1/2) & \sqrt{N^2 + N} & 0 \\ N + 1/2 & 0 & 0 & \sqrt{N^2 + N} \\ \sqrt{N^2 + N} & 0 & 0 & N + 1/2 \\ 0 & \sqrt{N^2 + N} & -(N + 1/2) & 0 \end{bmatrix}. \quad (12.107)$$

12.4 Gaussian channels

12.4.1 Open bosonic systems

In this subsection, we consider a class of quantum channels that arise naturally from the interaction of a bosonic system with a bosonic environment. Let Z_A, Z_B be symplectic spaces describing the input and output of the channel, and let Z_D, Z_E be their corresponding environments, so that

$$Z_A \oplus Z_D = Z_B \oplus Z_E = Z, \quad (12.108)$$

and let $W_A(z_A), \dots$ be the Weyl operators in the Hilbert spaces \mathcal{H}_A, \dots of the corresponding bosonic systems. Assume that the composite system in the Hilbert space

$$\mathcal{H}_A \otimes \mathcal{H}_D = \mathcal{H}_B \otimes \mathcal{H}_E = \mathcal{H} \quad (12.109)$$

is prepared in the initial state $S_A \otimes S_D$ and evolves according to the unitary operator U_T , which corresponds to a symplectic transformation T in Z by (12.64). In accordance with the direct sum decompositions (12.108), T can be written in block matrix form

$$T = \begin{bmatrix} K & L \\ K_D & L_D \end{bmatrix}, \quad (12.110)$$

where $K : Z_B \rightarrow Z_A, L : Z_E \rightarrow Z_A, K_D : Z_B \rightarrow Z_D, L_D : Z_E \rightarrow Z_D$.

The characteristic function of the system state after the interaction described by this unitary transformation U_T is

$$\begin{aligned} \varphi_B(z_B) &= \text{Tr } U_T (S_A \otimes S_D) U_T^* (W_B(z_B) \otimes I_E) \\ &= \text{Tr } (S_A \otimes S_D) U_T^* (W_B(z_B) \otimes W_E(0)) U_T \\ &= \text{Tr } (S_A \otimes S_D) (W_A(Kz_B) \otimes W_D(K_D z_B)) \\ &= \varphi_A(Kz_B) \cdot \varphi_D(K_D z_B), \end{aligned}$$

where $\varphi_D(z_D) = \text{Tr } S_D W_D(z_D)$ is the characteristic function of the initial state of the environment. This transformation can be written in the form

$$\varphi_B(z_B) = \varphi_A(Kz_B) \cdot f(z_B), \quad (12.111)$$

where

$$f(z_B) = \varphi_D(K_D z_B). \quad (12.112)$$

If the initial state of the environment is Gaussian, with parameters (m_D, α_D) , f has the form of the Gaussian characteristic function

$$f(z_B) = \exp \left[i l(z_B) - \frac{1}{2} \alpha(z_B, z_B) \right], \quad (12.113)$$

where $l(z_B) = m_D(K_D z_B)$ and

$$\alpha(z_B, z'_B) = \alpha_D(K_D z_B, K_D z'_B). \quad (12.114)$$

Transformation of states according to the formula (12.111) is a concatenation of the unitary evolution and partial trace, which are channels in the sense of Definition 11.16, and hence defines a quantum channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$.

Definition 12.29. A channel that transforms states according to the rule (12.111) is called *linear bosonic*. If, additionally, f is a Gaussian characteristic function (12.113), where l is a linear form and α is an inner product on Z_B , the channel is called *Gaussian*, with parameters (K, l, α) .

Let R_A, \dots be the vectors of canonical observables in \mathcal{H}_A, \dots , with the commutation matrices Δ_A, \dots . From the tensor product decomposition (12.109), we have two different splits of the set of canonical observables of the composite system:

$$[R_B R_E] = [R_A R_D].$$

According to (12.65), the action of the operator U_T is described by the linear transformation

$$[R'_B R'_E] \equiv U_T^* [R_B R_E] U_T = [R_A R_D] T, \quad (12.115)$$

where T is the block matrix (12.110) and, to simplify the notation, we write R_A, \dots instead of $R_A \otimes I_D, \dots$. In particular,

$$R'_B = R_A K + R_D K_D. \quad (12.116)$$

The commutation and the covariance matrices of the canonical observables R'_B at the output of the channel are given by the expression

$$\alpha_B + \frac{i}{2} \Delta_B = \text{Tr} (R'_B)^\top S R'_B,$$

where $S = S_A \otimes S_D$ and S_A, S_D are density operators in \mathcal{H}_A and \mathcal{H}_D , with the covariance matrices α_A and α_D , respectively. By using (12.116), we obtain

$$\Delta_B = K^\top \Delta_A K + K_D^\top \Delta_D K_D, \quad (12.117)$$

$$\alpha_B = K^\top \alpha_A K + K_D^\top \alpha_D K_D. \quad (12.118)$$

Note that in the Gaussian case $\alpha = K_D^\top \alpha_D K_D$ is the matrix of the corresponding form that enters the expression for the function (12.113) defining the channel. Since $\alpha_D \geq \pm \frac{i}{2} \Delta_D$, the relation (12.117) implies the inequality

$$\alpha \geq \pm \frac{i}{2} \left[\Delta_B - K^\top \Delta_A K \right], \quad (12.119)$$

which will play a key role in what follows.

Equation (12.116) is the “input-output relation” describing the channel in the way closest to the classical description in terms of “signal plus Gaussian noise”. Here, R_A is the quantum signal and $N_D = R_D K_D$ is the Gaussian noise with the covariance matrix $\alpha = K_D^\top \alpha_D K_D$. Note that the noise can be partially quantum and partially classical, because of the possible degeneracy of the commutator matrix $\Delta_B - K^\top \Delta_A K$.

The final state of the environment is defined by a similar equation for the canonical observables R_E after the interaction:

$$R'_E = R_A L + R_D L_D.$$

The weakly complementary channel $\Phi_E : S_A \rightarrow S_E$ (see Section 6.6) is defined by the relation

$$\varphi_E(z_E) = \varphi_A(Lz_E) \cdot \varphi_D(LDz_E),$$

and is also linear bosonic, and it is Gaussian if the state S_D is Gaussian. If the initial state of the environment is pure, $S_D = |\psi_D\rangle\langle\psi_D|$, the channel Φ_E is complementary to Φ , with the Stinespring isometry given by formula $V = U_T|\psi_D\rangle$.

Theorem 12.30. *Inequality (12.119) is a necessary and sufficient condition for a set of parameters (K, l, α) to define a Gaussian channel.*

Proof. The necessity of condition (12.119) was established before. Let us now show that under this condition one can construct the symplectic transformation T defining the dynamics of the open bosonic system which gives rise to the required Gaussian channel. In doing this, we can always restrict ourselves to the centered case ($l = 0$) by performing a displacement operation (12.81).

First of all, notice that, for the given operators K, K_D satisfying (12.117), one can always extend the transformation (12.116) to a symplectic transformation T . Indeed, relation (12.117) means that the transformation $z_B \rightarrow [Kz_B, K_Dz_B]^T$ is a symplectic embedding of Z_B into $Z_A \oplus Z_D = Z$, i.e. the columns of the matrix $[K, K_D]^T$ form a system that can be complemented to a symplectic basis in Z , which forms the matrix T . Thus, the problem reduces to the following. For given K, α , satisfying the inequality (12.119), find a symplectic space (Z_D, Δ_D) , an operator $K_D : Z_B \rightarrow Z_D$, and an inner product in Z_D , defined by symmetric matrix $\alpha_D \geq \pm \frac{i}{2}\Delta_D$, such that

$$K_D^\top \Delta_D K_D = \Delta_B - K^\top \Delta_A K \equiv \tilde{\Delta}_D; \quad (12.120)$$

$$K_D^\top \alpha_D K_D = \alpha. \quad (12.121)$$

Assume first that α is nondegenerate. Consider the Euclidean space (Z_B, α) with the skew-symmetric form $\tilde{\Delta}_D$. Let $2n = \dim Z_B$. By the assumption, $\alpha \geq \pm \frac{i}{2}\tilde{\Delta}_D$. Consider the skew-symmetric operator S , defined by the relation

$$\tilde{\Delta}_D(z_B, z'_B) = \alpha(z_B, Sz'_B); \quad z_B, z'_B \in Z_B.$$

According to a theorem from linear algebra, there is an orthonormal basis $\tilde{e}_j; j = 1, \dots, 2n-l; \tilde{h}_j; j = 1, \dots, l$ in (Z_B, α) such that

$$S\tilde{e}_j = s_j \tilde{h}_j, S\tilde{h}_j = -s_j \tilde{e}_j; \quad j = 1, \dots, l; \quad S\tilde{e}_j = 0; \quad j = l+1, \dots, 2n-l.$$

The condition $\alpha \geq \pm \frac{i}{2}\tilde{\Delta}_D$ implies $I - \frac{i}{2}S \geq 0$, hence $0 < s_j \leq 2$.

Let (Z_D, Δ_D) be the standard symplectic space of dimensionality $2l+2(2n-2l) = 2(2n-l)$, with the basis $e_j, h_j; j = 1, \dots, 2n-l$. Define α_D as a form with the diagonal matrix

$$\begin{aligned} \alpha_D e_j &= s_j^{-1} e_j, & \alpha_D h_j &= s_j^{-1} h_j; & j &= 1, \dots, l; \\ \alpha_D e_j &= e_j, & \alpha_D h_j &= \frac{1}{4} h_j; & j &= l+1, \dots, 2n-l. \end{aligned}$$

In this case, $\alpha_D \geq \pm \frac{i}{2} \Delta_D$. Define the operator $K_D : Z_B \rightarrow Z_D$ by the formula

$$\begin{aligned} K_D \tilde{e}_j &= \sqrt{s_j} e_j, & K_D \tilde{h}_j &= \sqrt{s_j} h_j; & j &= 1, \dots, l; \\ K_D \tilde{e}_j &= e_j, & & & j &= l+1, \dots, 2n-l. \end{aligned} \quad (12.122)$$

Now, relations (12.120) and (12.121) follow from the consideration of the values of the corresponding quadratic forms on the basis vectors \tilde{e}_j ; $j = 1, \dots, 2n-l$; \tilde{h}_j ; $j = 1, \dots, l$ in Z_B . Note that, by construction, α_D is the covariance matrix of a pure state if and only if $s_j = 2$; $j = 1, \dots, l$.

If, on the other hand, α vanishes on a nontrivial subspace of $Z_0 \subset Z_B$, then $\tilde{\Delta}_D$ also vanishes on Z_0 due to the condition $\alpha \geq \pm \frac{i}{2} \tilde{\Delta}_D$. Therefore, the vectors \tilde{e}_j ; $j = l+1, \dots, 2n-l$ can be chosen in such a way that part of them will form a basis in Z_0 . In this case, definition (12.122) can be modified by requiring $K_D \tilde{e}_j = 0$ for $\tilde{e}_j \in Z_0$, and relations (12.120) and (12.121) will be satisfied. \square

12.4.2 Gaussian channels: basic properties

Formula (12.111) is equivalent to the following equation for the dual channel

$$\Phi^*[W_B(z_B)] = W(K_B z_B) f(z_B), \quad (12.123)$$

which shows that the Weyl operators are mapped into Weyl operators, up to a factor. In the case of a Gaussian channel

$$\Phi^*[W_B(z_B)] = W(K_B z_B) \exp \left[i l(z_B) - \frac{1}{2} \alpha(z_B, z_B) \right]. \quad (12.124)$$

The following statement shows that Gaussian channels can be defined abstractly by relation (12.124) without reference to the picture of interaction with the environment considered in the previous section.

Proposition 12.31. *Condition (12.119) is necessary and sufficient for the mapping (12.124) to be completely positive.*

Proof. Sufficiency was established in Theorem 12.30. To prove necessity, note that complete positivity of the mapping (12.123) implies nonnegative definiteness of the matrices with operator elements

$$\begin{aligned} & W(K z_s) \Phi^*[W(z_s)^* W(z_r)] W(K z_r)^* \\ &= f(z_r - z_s) \exp \frac{i}{2} [\Delta(z_r, z_s) - \Delta(K^\top z_r, K^\top z_s)], \end{aligned} \quad (12.125)$$

where z_1, \dots, z_n is an arbitrary finite subset of Z . In the Gaussian case (12.124) this is equivalent to nonnegative definiteness of Hermitian matrices with elements

$$\exp \left\{ \alpha(z_r, z_s) - \frac{i}{2} \Delta(z_r, z_s) + \frac{i}{2} \Delta(K z_r, K z_s) \right\}, \quad (12.126)$$

and the condition (12.119) follows from Lemma 12.18. \square

We will make use of the following properties of Gaussian channels.

- i. A Gaussian channel transforms Gaussian states into Gaussian states. The parameters of the input state are transformed according to the relations

$$\begin{aligned} m_B &= m_A K + l, \\ \alpha_B &= K^\top \alpha_A K + \alpha, \end{aligned} \quad (12.127)$$

compare with relation (12.118).

- ii. The dual of a linear bosonic channel transforms any polynomial in the canonical observables R_B into a polynomial in R_A of the same order, provided the function f has derivatives of sufficiently high order. This property follows by differentiating relation (12.123) in the point $z_B = 0$.
- iii. A concatenation of Gaussian channels is a Gaussian channel. Indeed, let Φ_j ; $j = 1, 2$, be two Gaussian channels with parameters K_j, l_j, α_j . In this case, applying the definition (12.124), we obtain the Gaussian channel $\Phi_2 \circ \Phi_1$ with parameters

$$\begin{aligned} K &= K_1 K_2, \\ l &= K_2^\top l_1 + l_2, \\ \alpha &= K_2^\top \alpha_1 K_2 + \alpha_2. \end{aligned} \quad (12.128)$$

- iv. Any linear bosonic, in particular Gaussian, channel is covariant in the sense that

$$\Phi[W(z)^* S W(z)] = W(K'z)^* \Phi[S] W(K'z), \quad (12.129)$$

where $K' = \Delta^{-1} K^\top \Delta$ is the symplectic adjoint of K .

12.4.3 Gaussian observables

Assume that we have two symplectic spaces Z_A, Z_B with the corresponding Weyl systems in Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$. Let M be an observable in \mathcal{H}_A with the outcome set Z_B , given by the probability operator-valued measure $M(d^{2n}z)$. In what follows, we occasionally skip the index B so that $Z = Z_B$ etc. The observable is completely determined by the *operator characteristic function*

$$\phi_M(w) = \int e^{-i\Delta(w,z)} M(d^{2n}z).$$

Note the following apparent property of the operator characteristic function. For any choice of a finite subset $\{w_j\} \subset Z_{\mathbb{R}}$, the block matrix with operator entries $\phi(w_j - w_k)$ is nonnegative definite.

An observable M will be called *Gaussian* if its operator characteristic function has the form

$$\phi_M(w) = W_A(Kw) \exp\left(-\frac{1}{2}\alpha(w, w)\right) = \exp\left(iR_A Kw - \frac{1}{2}\alpha(w, w)\right), \quad (12.130)$$

where $K : Z_B \rightarrow Z_A$ is a linear operator and α is a bilinear form on Z_B . The pair (K, α) defines the parameters of the Gaussian observable.

Theorem 12.32. *A necessary and sufficient condition for relation (12.130) to define an observable is the matrix inequality*

$$\alpha \geq \pm \frac{i}{2} K^\top \Delta_A K. \quad (12.131)$$

Proof. For an operator function given by (12.130),

$$\phi_M(w_j - w_k) = C_k^* C_j \exp\left[\alpha(w_j, w_k) - \frac{i}{2} \Delta(Kw_j, Kw_k)\right],$$

where $C_j = W_A(Kw_j) \exp\left[-\frac{1}{2}\alpha(w_j, w_j)\right]$, and nonnegative definiteness of matrices with scalar entries $\exp\left[\alpha(w_j, w_k) - \frac{i}{2} \Delta(Kw_j, Kw_k)\right]$ implies inequality (12.131), according to Lemma 12.18.

Sufficiency of the condition (12.131) will be established by a direct construction of the Naimark extension of an observable M in the spirit of Corollary 3.8.

Proposition 12.33. *Assume condition (12.131). Then one can find a bosonic system in the space \mathcal{H}_C with canonical observables R_C , for which $\mathcal{H}_B \subseteq \mathcal{H}_A \otimes \mathcal{H}_C$, and the Gaussian state $S_C \in \mathfrak{S}(\mathcal{H}_C)$, such that*

$$M(U) = \text{Tr}_C (I_A \otimes S_C) E_{AC}(U), \quad U \subseteq Z_B, \quad (12.132)$$

where E_{AC} is the sharp observable in the space $\mathcal{H}_A \otimes \mathcal{H}_C$, given by the spectral measure of commuting self-adjoint operators

$$X_B = (R_A K + R_C K_C) \Delta_B^{-1}, \quad (12.133)$$

(here we again abbreviate the notations as R_A instead of $R_A \otimes I_C$ etc.) where $K_C : Z_B \rightarrow Z_C$ is an operator such that

$$K_C^\top \Delta_C K_C = -K^\top \Delta_A K. \quad (12.134)$$

Proof. Condition (12.134) means that $K_C^\top \Delta_C K_C + K^\top \Delta_A K = 0$, that is, commutativity of the operators defined by (12.133). By adapting the proof of Theorem 12.30, we obtain a symplectic space (Z_C, Δ_C) , an operator $K_C : Z_B \rightarrow Z_C$, and an inner

product in Z_C , given by symmetric matrix $\alpha_C \geq \pm \frac{i}{2} \Delta_C$ such that (12.134) holds along with

$$K_C^\top \alpha_C K_C = \alpha.$$

Then the characteristic function of the observable E_{AC} is

$$\begin{aligned} \phi_{E_{AC}}(w) &= \int e^{-i\Delta(w,z)} E_{AC}(d^{2n}z) \\ &= \exp(iX_B \Delta_B w) = \exp i(R_A K + R_C K_C) w \\ &= W_A(Kw) W_C(K_C w), \end{aligned}$$

whence

$$\begin{aligned} \text{Tr}_C(I_A \otimes S_C) \phi_{E_{AC}}(w) &= W_A(Kw) \exp\left(-\frac{1}{2} \alpha_C(K_C w, K_C w)\right) \\ &= W_A(Kw) \exp\left(-\frac{1}{2} \alpha(w, w)\right) = \phi_M(w), \end{aligned}$$

and (12.132) follows. \square

An observable M is sharp if and only if $\alpha = 0$, in which case it is the spectral measure of the commuting self-adjoint operators $R_A K$.

Example 12.34. If $R = [q \ p]$ are the canonical observables of one mode A , and $R_C = [q_C \ p_C]$ those of another mode C , the operators $X_B = [q' \ p']$, defined as

$$q' = q - q_C, \quad p' = p + p_C;$$

are commuting essentially self-adjoint. If S_C is an elementary Gaussian state (12.79) with parameters $(0, \alpha)$, we have an extension, in the sense of Proposition 12.33, of the Gaussian observable M with parameters (Δ, α) . In quantum optics, the observable M describes the statistics of optical heterodyning.

On the other hand, a single self-adjoint operator $X = k_q q + k_p p$, where k_q, k_p are real, describes the sharp Gaussian observable with parameters $([k_q \ k_p]^\top, 0)$ corresponding to optical homodyning. See, e.g. [38], [107] for more detail.

12.4.4 Gaussian entanglement-breaking channels

Theorem 12.35. A Gaussian channel Φ with parameters $(K, 0, \alpha)$ is entanglement-breaking if and only if α admits the decomposition

$$\alpha = \alpha_A + \alpha_B, \quad \text{where } \alpha_A \geq \pm \frac{i}{2} K^\top \Delta_A K, \quad \alpha_B \geq \pm \frac{i}{2} \Delta_B. \quad (12.135)$$

In this case, Φ has the representation

$$\Phi[S] = \int_{Z_B} W_B(z) S_B W_B(z)^* \mu_S(d^{2n}z), \quad (12.136)$$

where S_B is the Gaussian state with parameters $(0, \alpha_B)$, and $\mu_S(U) = \text{Tr } S M_A(U)$, $U \subseteq Z_B$ is the probability distribution of the Gaussian observable M_A with characteristic function

$$\phi_{M_A}(w) = W_A(Kw) \exp\left(-\frac{1}{2}\alpha_A(w, w)\right). \quad (12.137)$$

Proof. First, assume that α admits the decomposition (12.135) and consider the channel defined by (12.136). We have to show that

$$\Phi^*[W_B(w)] = W_A(Kw) \exp\left[-\frac{1}{2}\alpha(w, w)\right]. \quad (12.138)$$

Indeed, for an arbitrary state S

$$\begin{aligned} \text{Tr } S \Phi^*[W_B(w)] &= \text{Tr } \Phi[S] W_B(w) = \int_{Z_B} \text{Tr } W_B(z) S_B W_B(z)^* W_B(w) \mu_S(d^{2n}z) \\ &= \int_{Z_B} \text{Tr } S_B W_B(z)^* W_B(w) W_B(z) \mu_S(d^{2n}z) \\ &= \text{Tr } S_B W_B(w) \int_{Z_B} \exp[-i \Delta(w, z)] \mu_S(d^{2n}z) \\ &= \exp\left[-\frac{1}{2}\alpha_B(w, w)\right] \text{Tr } S \phi_{M_A}(w) \\ &= \text{Tr } S W_A(Kw) \exp\left[-\frac{1}{2}\alpha_B(w, w) - \frac{1}{2}\alpha_A(w, w)\right], \end{aligned} \quad (12.139)$$

whence (12.138) follows.

Conversely, let the channel Φ be Gaussian and entanglement-breaking. We will use the Gaussian version of the proof of Theorem 11.31. Fix a nondegenerate Gaussian state σ_A in $\mathfrak{S}(\mathcal{H}_A)$ and let $\{|e_j\rangle\}_{j=1}^{+\infty}$ be the basis of the eigenvectors of σ_A , with the corresponding (positive) eigenvalues $\{\lambda_j\}_{j=1}^{+\infty}$. Consider the unit vector

$$|\Omega\rangle = \sum_{j=1}^{+\infty} \sqrt{\lambda_j} |e_j\rangle \otimes |e_j\rangle$$

in the space $\mathcal{H}_A \otimes \mathcal{H}_A$. Then $|\Omega\rangle\langle\Omega|$ is the Gaussian purification of σ_A . Since Φ is entanglement-breaking, the Gaussian state

$$\sigma_{AB} = (\text{Id}_A \otimes \Phi)[|\Omega\rangle\langle\Omega|] \quad (12.140)$$

in $\mathfrak{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is separable. This implies representation (12.102), i.e.

$$\sigma_{AB} = \int_{Z_A} \int_{Z_B} W_A(z_A) S_A W_A(z_A)^* \otimes W_B(z_B) S_B W_B(z_B)^* P(d^{2m}z_A d^{2n}z_B), \quad (12.141)$$

where S_A, S_B are Gaussian states with covariance matrices α_A, α_B , and P is a Gaussian probability distribution.

Taking the partial trace with respect to B , this produces

$$\begin{aligned}\sigma_A &= \text{Tr}_B(\text{Id}_A \otimes \Phi)[|\Omega\rangle\langle\Omega|] \\ &= \int_{Z_A} \int_{Z_B} W_A(z_A) S_A W_A(z_A)^* P(d^{2m} z_A d^{2n} z_B) \\ &= \int_{Z_A} \int_{Z_B} \overline{W_A(z_A) S_A W_A(z_A)^*} P(d^{2m} z_A d^{2n} z_B),\end{aligned}$$

where the bar means complex conjugation in the basis of eigenvectors of σ_A . Now, similar to (11.54), we conclude that the relation

$$M_A(U) = \sigma_A^{-1/2} \left[\int_{Z_A} \int_U \overline{W_A(z_A) S_A W_A(z_A)^*} P(d^{2m} z_A d^{2n} z_B) \right] \sigma_A^{-1/2}$$

defines a bounded positive operator, such that $M_A(U) \leq M_A(Z_B) = I_A$. It is easy to see that $M_A(U)$ is a probability operator valued measure on Borel subsets $U \subseteq Z_B$.

Let us show that representation (12.136) holds for channel Φ with these M_A and S_B . Consider the entanglement-breaking channel

$$\hat{\Phi}[S] = \int_{Z_B} W_B(z) S_B W_B(z)^* \mu_S(d^{2n} z).$$

where $\mu_S(U) = \text{Tr } S M_A(U)$; $U \subseteq Z_B$. To prove $\Phi = \hat{\Phi}$, it is sufficient to show that $\hat{\Phi}[|e_j\rangle\langle e_k|] = \Phi[|e_j\rangle\langle e_k|]$ for all j, k . But

$$\begin{aligned}\hat{\Phi}[|e_j\rangle\langle e_k|] &= \int_{Z_B} W_B(z) S_B W_B(z)^* \langle e_k | M_A(d^{2n} z) | e_j \rangle \\ &= \lambda_j^{-1/2} \lambda_k^{-1/2} \int_{Z_A} \int_{Z_B} \left[\langle e_j | W_A(z_A) S_A W_A(z_A)^* | e_k \rangle \right. \\ &\quad \left. \cdot W_B(z_B) S_B W_B(z_B)^* P(d^{2m} z_A d^{2n} z_B) \right] \\ &= \lambda_j^{-1/2} \lambda_k^{-1/2} \text{Tr}_A(|e_k\rangle\langle e_j| \otimes I_B) \sigma_{AB} = \Phi[|e_j\rangle\langle e_k|],\end{aligned}$$

according to (12.140), (12.141).

It remains for us to show that M_A is Gaussian observable, with the characteristic function (12.137), where $\alpha_A = \alpha - \alpha_B$, and α_B is the covariance matrix of the state S_B ; without loss of generality, we can assume that its mean is zero. Indeed, the density

operator S_B can be modified with the help of the Weyl operators to have zero mean, resulting only in a change in the probability distribution (12.141) which, however, will remain Gaussian. But, from (12.139),

$$\Phi^*[W_B(w)] = \exp\left[-\frac{1}{2}\alpha_B(w, w)\right] \phi_{M_A}(w)$$

for any channel Φ with the representation (12.136), whence, taking into account (12.138), we indeed obtain (12.137), with $\alpha_A = \alpha - \alpha_B$. \square

A necessary condition for the decomposability (12.135) and hence for the channel to be entanglement-breaking is

$$\alpha \geq \frac{i}{2} \left(\Delta_B \pm K^T \Delta_A K \right). \quad (12.142)$$

In general, this condition means that for any input Gaussian state of the channel $\text{Id}_A \otimes \Phi$, the output has a positive partial transpose. Indeed, this channel transforms the covariance matrix of the input state according to the rule

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \rightarrow \begin{bmatrix} I & 0 \\ 0 & K^T \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & K \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \alpha \end{bmatrix} \equiv \alpha_{AB}.$$

The right hand side, representing the covariance matrix of the output state, satisfies

$$\begin{aligned} \alpha_{AB} &\geq \frac{i}{2} \begin{bmatrix} I & 0 \\ 0 & K^T \end{bmatrix} \begin{bmatrix} \Delta_A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & K \end{bmatrix} + \frac{i}{2} \begin{bmatrix} 0 & 0 \\ 0 & \pm \Delta_B - K^T \Delta_A K \end{bmatrix} \\ &= \frac{i}{2} \begin{bmatrix} \Delta_A & 0 \\ 0 & \pm \Delta_B \end{bmatrix}, \end{aligned}$$

where in the estimate of the second term we used (12.142) with its transpose. However, this is equivalent to condition (12.103), which is necessary and sufficient for the output state to have a positive partial transpose. Thus, condition (12.142) characterizes the class of Gaussian PPT channels, which is in general larger than the class of entanglement-breaking channels.

The condition of the theorem is automatically satisfied in the special case where

$$K^T \Delta_A K = 0.$$

In this case, the operators $R_A K$ commute and hence M_A is a sharp observable given by their spectral measure, and the probability distribution $\mu_S(d^{2n}z)$ can be arbitrarily sharply peaked around any point z by an appropriate choice of the state S . Hence, in this case, it is natural to identify Φ as a c-q (classical-quantum) channel determined by the family of states $z \rightarrow W(z)S_B W(z)^*$.

12.5 The capacities of Gaussian channels

12.5.1 Maximization of the mutual information

We start with the classical entanglement-assisted capacity, which can be computed most effectively in the Gaussian case. When the state S and the channel Φ are Gaussian, the quantities $H(S)$, $H(\Phi[S])$, $H(S, \Phi)$ and $I(S, \Phi)$ can be found by using formulas (12.97), (12.118), (12.105). Namely, $H(\Phi[S])$ is given by formula (12.97), with α replaced by $\alpha' = \alpha_B$ computed via (12.118). By purifying the input state and using formula (7.49), we obtain

$$H(S, \Phi) = \frac{1}{2} \text{Sp } g \left(\text{abs}(\Delta_{12}^{-1} \alpha'_{12}) - \frac{I}{2} \right),$$

where the matrix

$$\alpha'_{12} = \begin{bmatrix} \alpha' & K\beta \\ \beta^\top K^\top & \alpha \end{bmatrix} \quad (12.143)$$

is computed by inserting R'_B , given by (12.116), into

$$\alpha'_{12} - \frac{i}{2} \Delta'_{12} = \text{Tr} [R'_B R_2]^\top S [R'_B R_2],$$

where R_2 are the (unchanged) canonical observables of the reference system. Alternatively, if an explicit description of the complementary channel $\tilde{\Phi}$ is available, the entropy exchange can be calculated as its output entropy $H(\tilde{\Phi}[S])$.

The following result greatly simplifies the computation of the entanglement-assisted capacity of the Gaussian channels, by reducing to the case of Gaussian input states.

Theorem 12.36. *Let Φ be a Gaussian channel. The maximum of the mutual information $I(S, \Phi)$ over the set of states $\Sigma(m, \alpha)$ with given first and second moments is achieved on a Gaussian state.*

Proof. By using the representation in (7.53) for the quantum mutual information, we can write

$$I(S, \Phi) = H(B|E) + H(B),$$

where B is the output of the channel and E is the environment. For a Gaussian channel, the first and second moments are transformed in the same way for all states in $\Sigma(m, \alpha)$. By using Lemma 12.25, we have

$$H(B) = H(\Phi[S]) \leq H(\widehat{\Phi[S]}) = H(\Phi[\tilde{S}]) = H(\tilde{B}),$$

where $\widehat{\Phi[S]}$ (respectively, \tilde{S}) is the Gaussian state with the same first and second moments as $\Phi[S]$ (resp. S). The channel $S \rightarrow S_{BE} = VSV^*$, where V is the

Stinespring isometry for Φ , is also Gaussian, since it is implemented by a unitary operator U_T with Gaussian state of environment, namely

$$\text{Tr } S_{BE} W(z_B) \otimes W(\tilde{z}_E) = \phi_A(Kz_B + Lz_E) \phi_D(K_D z_B + L_D z_E). \quad (12.144)$$

Therefore, a similar argument, based on Lemma 12.26, implies that $H(B|E) \leq H(\tilde{B}|\tilde{E})$. Thus $I(S, \Phi) \leq I(\tilde{S}, \Phi)$. \square

This theorem implies that if the maximum of $I(S, \Phi)$ over a set of density operators defined by an arbitrary constraint on the first and second moments is achieved, this is achieved on a Gaussian density operator. Now, consider the energy constraint

$$\text{Tr } SF \leq E, \quad (12.145)$$

where $F = R\epsilon R^\top$ is the quadratic operator with positive, nondegenerate energy matrix ϵ . Notice, that

$$\text{Tr } SF = \text{Sp}(\epsilon\alpha_A) + m_A \epsilon m_A^\top,$$

where m_A is the mean vector and α_A is the covariance matrix of the state S . The expression for the entanglement-assisted classical capacity $C_{ea}(\Phi)$ given by Proposition 11.26 is the maximum of the quantum mutual information $I(S, \Phi)$ over all states S satisfying the constraint (12.145). The set of conditions for this proposition is satisfied for the operator $F = R\epsilon R^\top$. Indeed, by Theorem 12.22, the operator F satisfies condition (11.9). Taking

$$\tilde{F} = c[RR^\top - (\text{Sp}\alpha_E K_E^\top K_E)I],$$

we have $\Phi^*[\tilde{F}] = cRK^\top K R^\top$ and we can always choose positive c such that $\Phi^*[\tilde{F}] \leq F$. Moreover, \tilde{F} satisfies condition (11.9). Therefore, the maximum of the quantity $I(S, \Phi)$ is achieved and formula (11.41) holds. Since the energy constraint can be expressed in terms of m_A, α_A , the maximal value $C_{ea}(\Phi, F, E)$ is achieved on the Gaussian state.

Further reduction of the maximization problem can be obtained by invoking gauge covariance.

12.5.2 Gauge-covariant channels

Consider a channel $\Phi : \mathfrak{T}(\mathcal{H}_A) \rightarrow \mathfrak{T}(\mathcal{H}_B)$. Assume that in the spaces Z_A, Z_B some operators of complex structure J_A, J_B are fixed and let G_A, G_B be the operators generating the unitary groups of gauge transformations in $\mathcal{H}_A, \mathcal{H}_B$, according to formula (12.73). The channel is called *gauge covariant*, if

$$\Phi[e^{i\varphi G_A} S e^{-i\varphi G_A}] = e^{i\varphi G_B} \Phi[S] e^{-i\varphi G_B} \quad (12.146)$$

for all input states S and all $\varphi \in [0, 2\pi]$. In the case of a linear bosonic channel Φ , due to (12.73) and (12.111), this is equivalent to the following

$$\phi(e^{-\varphi J_A} K z_B) f(z_B) = \phi(K e^{-\varphi J_B} z_B) f(e^{-\varphi J_B} z_B),$$

where $\phi(z_A)$ is the characteristic function of the state S . For the Gaussian channel with parameters $(K, 0, \alpha)$ this reduces to

$$K J_B - J_A K = 0, \quad [\Delta_B^{-1} \alpha, J_B] = 0. \quad (12.147)$$

Thus, a natural choice of the complex structure in Z_B is given by any J_B , commuting with the operator $\Delta_B^{-1} \alpha$. Existence of such a complex structure is proved similar to (12.86), with the difference that the matrix α can be degenerate.

In optimization problems with bounded mean energy, a natural complex structure in Z_A is determined by the energy operator $F = R \epsilon R^\top$, namely, J_A is the operator of complex structure in Z_A , commuting with the operator $\epsilon \Delta_A$, so that

$$J_A \epsilon + \epsilon J_A^\top = 0. \quad (12.148)$$

In the case where F is the usual Hamiltonian of the oscillator system, the action of J_A reduces to multiplication by i in the complexification associated with creation-annihilation operators (see (12.63)).

Now, let the Gaussian channel Φ be gauge covariant with respect to these natural complex structures. In this case,

$$I(e^{i\varphi G_A} S e^{-i\varphi G_A}, \Phi) \equiv I(S, \Phi),$$

which follows from the similar property of the three terms that comprise $I(S, \Phi)$. For $H(S)$, this is simply a consequence of the unitary invariance of the entropy, for $H(\Phi[S])$ the covariance of the channel Φ is additionally used, and for $H(S, \Phi) = H(\tilde{\Phi}[S])$ this follows from the covariance of the complementary channel $\tilde{\Phi}$ (Exercise 6.38). Defining the average G_A -invariant state

$$\bar{S} = \frac{1}{2\pi} \int_0^{2\pi} e^{i\varphi G_A} S e^{-i\varphi G_A} d\varphi,$$

we have

$$\text{Tr } S F = \text{Sp}(\epsilon \alpha_A) + m_A \epsilon m_A^\top \geq \text{Sp}(\epsilon \alpha_A) = \text{Tr } \bar{S} F,$$

where m_A, α_A are the first and the second moments of the state S and the last equality follows because of (12.148):

$$\begin{aligned} \text{Tr } \bar{S} F &= \frac{1}{2\pi} \int_0^{2\pi} \text{Tr } e^{i\varphi G_A} S e^{-i\varphi G_A} (R \epsilon R^\top) d\varphi \\ &= \frac{1}{2\pi} \int_0^{2\pi} \text{Sp} \left(e^{\varphi J_A} \epsilon e^{\varphi J_A^\top} \alpha_A \right) d\varphi = \text{Sp}(\epsilon \alpha_A). \end{aligned}$$

Now, using the concavity of the mutual information $I(S, \Phi)$, we conclude that $I(S, \Phi) \leq I(\bar{S}, \Phi)$. Thus, the maximum of the quantum mutual information over the set of states with bounded mean energy $\text{Tr } SF \leq E$ is attained on a gauge invariant (G_A -invariant) Gaussian state. The first and second moments of such a state necessarily satisfy the relations $m_A = 0$, $[J_A, \Delta^{-1}\alpha_A] = 0$. Considering the Gaussian state with these first and second moments, we finally obtain the following corollary.

Corollary 12.37. *Let Φ be a Gaussian channel that is gauge-covariant with respect to the natural complex structures J_A, J_B (where J_A is associated with the energy operator). In this case, the maximum of the quantum mutual information over the set of states with bounded mean energy is attained on a gauge-invariant (G_A -invariant) Gaussian state.*

12.5.3 Maximization of the coherent information

Consider the coherent information

$$I_c(S, \Phi) = H(\Phi[S]) - H(S, \Phi).$$

Proposition 12.38. *Let Φ be a Gaussian channel that is degradable*

$$\tilde{\Phi} = \Gamma \circ \Phi \tag{12.149}$$

such that Γ is Gaussian channel. In this case, the quantum capacity

$$Q(\Phi) = \sup_{\tilde{S}} I_c(\tilde{S}, \Phi),$$

where the supremum is taken over Gaussian states \tilde{S} . If, in addition, the channels are gauge-covariant, then the supremum can be taken only over gauge-invariant (G_A -invariant) Gaussian states.

Proof. Assuming that the channel is degradable, i.e. (12.149) holds, where $\tilde{\Phi}$ is the complementary channel and Γ is some channel, we have (see Section 10.3.3)

$$I_c(S, \Phi) = H(E'|E) = H(S_{E'E}) - H(S_E), \tag{12.150}$$

where $S_{E'E} = V'S_B V'^*$ and $V' : \mathcal{H}_B \rightarrow \mathcal{H}_E \otimes \mathcal{H}_{E'}$ is the minimal Stinespring isometry for the channel Γ . Now if the channel Γ can be chosen Gaussian, the channel $S_B \rightarrow S_{E'E}$ is also Gaussian and the argument invoking Lemma 12.26 applies. Therefore, $I_c(S, \Phi) = H(E'|E) \leq H(\tilde{E}'|\tilde{E}) = I_c(\tilde{S}, \Phi)$. By using Proposi-

tion 10.27, it follows that the quantum capacity of the degradable channel is given by

$$Q(\Phi) = \sup_S I_c(S, \Phi) = \sup_{\tilde{S}} I_c(\tilde{S}, \Phi),$$

where the supremum is taken over Gaussian states \tilde{S} .

If, in addition, Φ is gauge covariant, the concavity of the conditional quantum entropy (12.150) implies that the supremum can be taken only over gauge-invariant Gaussian states. \square

12.5.4 The classical capacity: conjectures

It is natural to consider the classical capacity of a quantum Gaussian channel Φ under the additive input constraint (11.28) with

$$F^{(n)} = F \otimes I \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes I \otimes F,$$

where $F = R\epsilon R^\top$ is the quadratic energy operator with positive-definite matrix ϵ . However, finding the classical capacity $C(\Phi, F, E)$ in general depends on the solution of the problem of the *additivity of the constrained χ -capacity* (11.33) which is still open for the Gaussian channels with quadratic input constraints. In any case, the χ -capacity $C_\chi(\Phi, F, E)$, defined by relation (11.31), is a lower estimate for $C(\Phi, F, E)$ that coincides with $C(\Phi, F, E)$ in the case of additivity, e.g. for entanglement-breaking channels. A related open problem is the *additivity of the minimal output entropy* (8.33) in the class of Gaussian channels.

However, even the computation of $C_\chi(\Phi, F, E)$ for Gaussian channels remains, in general, an open problem (except for the case of c-q channels and a special case, which will be considered in Proposition 12.46). At least, in this situation an optimal generalized ensemble always exists, because the conditions of Corollary 11.24 for $F = R\epsilon R^\top$ are satisfied, as was shown in Section 12.5.1. Thus, the maximum of the quantity $\chi_\Phi(\pi)$ defined in (11.35) is achieved, and $C_\chi(\Phi, F, E)$ is given by relation (11.38).

Consider the following **hypothesis of Gaussian optimal ensembles**: *For a Gaussian channel Φ with the quadratic input energy constraint, the maximum of the quantity $\chi_\Phi(\pi)$ in expression (11.38) for $C_\chi(\Phi, F, E)$ is attained on the Gaussian generalized ensemble π that consists of generalized coherent states $W(z)S_0W(z)^*$, where S_0 is a pure Gaussian state, with a Gaussian probability distribution $P(d^{2n}z)$.*

For such an ensemble the covariance property (12.129) implies

$$H(\Phi[W(z)S_0W(z)^*]) = H(\Phi[S_0])$$

hence

$$\chi_\Phi(\pi) = H(\Phi[\tilde{S}_\pi]) - H(\Phi[S_0]), \quad (12.151)$$

which leads us to the **hypothesis of the Gaussian minimizer for the output entropy**: For a Gaussian channel Φ , the minimum of the output entropy is attained on a (pure) Gaussian state S_0 .

Consider the gauge transformations in \mathcal{H}_A that are associated with an operator of complex structure. Define their action on the generalized ensembles by the formula

$$\pi_\varphi(U) = \pi(\{S : e^{i\varphi G_A} S e^{-i\varphi G_A} \in U\}), \quad \varphi \in [0, 2\pi],$$

for Borel subsets $U \in \mathfrak{S}(\mathcal{H}_A)$. A generalized ensemble is *gauge-invariant*, if $\pi_\varphi \equiv \pi$. By using the concavity of the functional $\chi_\Phi(\pi)$ (Exercise 11.23) and arguing as in Section 12.5.2, one can show that if the Gaussian channel is gauge covariant with respect to the complex structure associated with the energy matrix ϵ , the maximum in (11.38) is attained on a gauge-invariant generalized ensemble. Again, it follows that the average state \bar{S}_π of such an optimal ensemble is gauge-invariant.

However, under the same assumption one can prove the following proposition, relating the two hypotheses for gauge-covariant channels.

Proposition 12.39. *Let a Gaussian channel Φ be gauge-covariant with respect to the complex structures J_A, J_B . Assume that the minimum of the output entropy is attained on a G_A -invariant Gaussian state S_0 . In this case, the hypothesis of optimal Gaussian ensembles is valid, and the optimal ensemble π can be chosen such that the output state $\bar{S}_B = \Phi[\bar{S}_\pi]$ is a G_B -invariant Gaussian state.*

Proof. It was already observed that an optimal ensemble π exists. Let $\bar{S}_B = \Phi[\bar{S}_\pi]$. Now, $\text{Tr } \bar{S}_\pi F \leq E$ and

$$\begin{aligned} C_\chi(\Phi, F, E) &= \chi_\Phi(\pi) = H(\bar{S}_B) - \int H(\Phi[S])\pi(dS) \\ &\leq H(\bar{S}_B) - \check{H}(\Phi) = H(\bar{S}_B) - H(\Phi[S_0]), \end{aligned} \quad (12.152)$$

by the assumption $H(\Phi[S_0]) = \check{H}(\Phi)$.

Consider a G_A -invariant state

$$\bar{S}_A = \frac{1}{2\pi} \int_0^{2\pi} e^{i\varphi G_A} \bar{S}_\pi e^{-i\varphi G_A} d\varphi,$$

then $\text{Tr } \bar{S}_A F = \text{Tr } \bar{S}_\pi F$. Now, let \tilde{S}_A be the Gaussian state with the same first and second moments as \bar{S}_A . In this case, again,

$$\text{Tr } \tilde{S}_A F = \text{Tr } \bar{S}_A F = \text{Tr } \bar{S}_\pi F. \quad (12.153)$$

Moreover, \tilde{S}_A is a G_A -invariant Gaussian state and $\tilde{S}_B = \Phi[\tilde{S}_A]$ is a G_B -invariant Gaussian state with the same first and second moments as $\Phi[\bar{S}_A]$, due to the Gaussianity of Φ . Moreover,

$$H(\tilde{S}_B) \geq H(\Phi[\tilde{S}_A]) \geq H(\Phi[S_\pi]) = H(\bar{S}_B). \quad (12.154)$$

Here, the first inequality follows from the principle of maximal entropy (Lemma 12.25) while the second follows from the concavity of the entropy. Since S_0 is a pure G_A -invariant Gaussian state, one has the decomposition (12.88) for the state

$$\tilde{S}_A = \int W(z) S_0 W(z)^* P(d^{2s} z).$$

Denote by $\tilde{\pi}$ the ensemble of generalized coherent states $W(z) S_0 W(z)^*$ with Gaussian distribution $P(d^{2s} z)$. Then

$$\tilde{S}_B = \Phi[\tilde{S}_A] = \int \Phi[W(z) S_0 W(z)^*] P(d^{2s} z) \equiv \int \Phi[S] \tilde{\pi}(dS)$$

and

$$\int H(\Phi[S]) \tilde{\pi}(dS) = H(\Phi[S_0]) = \check{H}(\Phi).$$

By relation (12.153), ensemble $\tilde{\pi}$ satisfies the energy constraint. Moreover,

$$\chi_\Phi(\tilde{\pi}) = H(\tilde{S}_B) - \int H(\Phi[S]) \tilde{\pi}(dS) = H(\tilde{S}_B) - H(\Phi[S_0]). \quad (12.155)$$

Bringing together relations (12.152), (12.154), and (12.155), we obtain $\chi_\Phi(\tilde{\pi}) \geq \chi_\Phi(\pi) = C_\chi(\Phi, F, E)$. Therefore, $\tilde{\pi}$ is the optimal Gaussian ensemble having the required properties. \square

The condition of Proposition 12.39 is satisfied in the special case where there exists a G_A -invariant Gaussian state S_0 such that $\Phi[S_0]$ is pure, so that the minimal output entropy $\check{H}(\Phi) = 0$. In this direction, we have the following result.

Proposition 12.40. *Let the Gaussian channel Φ be gauge-covariant with respect to the complex structures J_A, J_B . Assume that the parameters of the channel $(K, 0, \alpha)$ satisfy the following conditions.*

- i. *the matrix $\Delta_K = \Delta_B - K^\top \Delta_A K$ is nondegenerate*
- ii. *J_B is an operator of complex structure in the symplectic space (Z_B, Δ_K)*
- iii. *$\alpha = \frac{1}{2} \Delta_K J_B$*

Let S_0 be the pure G_A -invariant Gaussian state with parameters $(0, \alpha_A)$, where $\alpha_A = \frac{1}{2} \Delta_A J_A$. Then the hypothesis of optimal Gaussian ensembles is valid, with the optimal Gaussian ensemble $\tilde{\pi}$ described in the proof of Proposition 12.39.

Moreover, the additivity property (11.33) holds for this channel, and

$$C(\Phi, F, E) = C_\chi(\Phi, F, E) = \max_{S: \text{Tr } SF \leq E} H(\Phi[S]), \quad (12.156)$$

where the maximum is attained on a G_A -invariant Gaussian state.

Proof. Recall that condition ii. requires that $\Delta_K J_B = -J_B^\top \Delta_K$ is a positive definite matrix. Moreover, statement iv. in Exercise 12.19 implies that $\alpha = \frac{1}{2} \Delta_K J_B$ is the covariance matrix of a pure Gaussian state over (Z_B, Δ_K) and hence satisfies $\alpha \geq \pm \frac{i}{2} \Delta_K$. But this is just the necessary and sufficient condition (12.119) for the triple $(K, 0, \alpha)$ to define a Gaussian channel. Hence, condition iii. is consistent. Let us show that $\Phi[S_0]$ is a pure Gaussian state with the parameters $(0, \alpha_B)$, where $\alpha_B = \frac{1}{2} \Delta_B J_B$, which implies $\check{H}(\Phi) = \Phi[S_0] = 0$ and hence the conclusion of Proposition 12.39.

According to (12.127), the covariance matrix of the output state $\Phi[S_0]$ is

$$\alpha_B = K^\top \alpha_A K + \alpha = \frac{1}{2} \left(K^\top \Delta_A J_A K + \Delta_K J_B \right).$$

The condition $J_A K = K J_B$ (see (12.147)) and the definition of Δ_K imply that this is equal to

$$\alpha_B = \frac{1}{2} \left(K^\top \Delta_A K J_B + \Delta_K J_B \right) = \frac{1}{2} \Delta_B J_B$$

which proves the first statement.

To prove (12.156), we first notice that, by an argument similar to the one used to deal with the maximum of the mutual quantum information in Section 12.5.1, the maximum in (12.156) is indeed attained on a G_A -invariant Gaussian state \tilde{S}_A . Now,

$$\max_{S: \text{Tr } S F \leq E} H(\Phi[S]) = H(\Phi[\tilde{S}_A]) = \chi_\Phi(\tilde{\pi}) = C_\chi(\Phi, F, E), \quad (12.157)$$

where $\tilde{\pi}$ is the Gaussian ensemble constructed from \tilde{S}_A , as in the proof of Proposition 12.39. Now consider the channel $\Phi^{\otimes n}$ for which also the minimal output entropy $\check{H}(\Phi^{\otimes n}) = 0$. Thus,

$$\begin{aligned} n C_\chi(\Phi, F, E) &\leq C_\chi(\Phi^{\otimes n}, F^{(n)}, nE) \\ &\leq \max_{S^{(n)}: \text{Tr } S^{(n)} F^{(n)} \leq nE} H(\Phi^{\otimes n}[S^{(n)}]) \\ &\leq n \max_{S: \text{Tr } S F \leq E} H(\Phi[S]). \end{aligned}$$

Here, the first and second inequalities follow from definition (11.31) of C_χ and the last one from Lemma 11.20. Combined with (12.157), this implies the additivity property $C_\chi(\Phi^{\otimes n}, F^{(n)}, nE) = n C_\chi(\Phi, F, E)$ and hence (12.156). \square

12.6 The case of one mode

12.6.1 Classification of Gaussian channels

In this section we are interested in the problem of reducing an arbitrary quantum Gaussian channel to the simplest “normal” form, which can be obtained by applying

suitable unitary operators, corresponding to symplectic transformations T_1, T_2 via the formula (12.64), to the input and the output of the channel :

$$\Phi^{*'}[W(z)] = U_{T_1}^* \Phi^* [U_{T_2}^* W(z) U_{T_2}] U_{T_1}$$

That is,

$$\Phi^{*'}[W(z)] = W(T_1 K T_2 z) f(T_2 z).$$

Here, we provide a complete solution of this problem in one mode, $s = 1$.

Let Z be the two-dimensional symplectic space, i.e. the real linear space of vectors $z = [x \ y]^T$ with the symplectic form

$$\Delta(z, z') = xy' - x'y. \quad (12.158)$$

A basis e, h in Z is symplectic if $\Delta(e, h) = 1$, i.e. if the area of the oriented parallelogram based on e, h is equal to 1. Recall that a linear transformation T in Z is symplectic if it maps a symplectic basis into a symplectic basis.

According to formula (12.124), a Gaussian channel is characterized by the parameters K, l, α satisfying

$$\alpha \geq \pm \frac{i}{2} \left[\Delta - K^T \Delta K \right]. \quad (12.159)$$

By way of a displacement transformation (12.52), we can always make $l = 0$, which we assume. Thus,

$$\Phi^*[W(z)] = W(Kz) \exp \left[-\frac{1}{2} z^T \alpha z \right]. \quad (12.160)$$

In the theorem below, we use the notation

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Theorem 12.41. *Let $\{e, h\}$ be a symplectic basis. Then depending on the value*

- | | |
|---|--------------------------------|
| A) $\Delta(Ke, Kh) = 0$; | B) $\Delta(Ke, Kh) = 1$; |
| C) $\Delta(Ke, Kh) = k^2 > 0, k \neq 1$ | D) $\Delta(Ke, Kh) = -k^2 < 0$ |

there are symplectic transformations T_1, T_2 , such that the channel $\Phi^{'}$ has the form (12.160), with*

$$\begin{aligned} A_1) \quad K &= 0; \\ \alpha &= \left(N_0 + \frac{1}{2} \right) I_2; \quad N_0 \geq 0; \end{aligned}$$

$$\begin{aligned}
A_2) \quad K &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}; \\
\alpha &= \left(N_0 + \frac{1}{2} \right) I_2 \\
B_1) \quad K &= I_2; \\
\alpha &= \frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}; \\
B_2) \quad K &= I_2; \\
\alpha &= N_c I_2; \quad N_c \geq 0; \\
C) \quad K &= k I_2; \quad k > 0, k \neq 1; \\
\alpha &= \left(N_c + \frac{|k^2 - 1|}{2} \right) I_2; \\
D) \quad K &= k \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \quad k > 0; \\
\alpha &= \left(N_c + \frac{(k^2 + 1)}{2} \right) I_2.
\end{aligned}$$

Apparently, in the cases $A)$, $B_2)$, $C)$, the channel is gauge-covariant with respect to the natural complex structure associated with multiplication by i in the complex plane x, y .

Proof.

$A)$ $\Delta(Ke, Kh) = 0$. In this case, $\Delta(Kz, Kz') \equiv 0$ and either $K = 0$ or K has rank one. Now, inequality (12.159) is just the condition on the covariance matrix of a quantum Gaussian state. As follows from Lemma 12.12, there is a symplectic transformation T_2 such that

$$\alpha(T_2 z, T_2 z) = \left(N_0 + \frac{1}{2} \right) (x^2 + y^2). \quad (12.161)$$

In the case where $K = 0$, we just have $A_1)$.

Let K have rank one. In this case, KT_2 has rank one and there is a vector e' such that $KT_2[x \ y]^T = (k_1 x + k_2 y) e'$. Now, there is a symplectic transformation T_1 such that $T_1 e' = [1 \ 0]^T$ and hence $T_1 K T_2 [x \ y]^T = [k_1 x + k_2 y \ 0]^T$. By making a rotation T'_2 that leaves α unchanged, we can transform this vector to $[k'_1 x \ 0]^T$ with $k'_1 \neq 0$, and then, by a symplectic scaling, (squeezing) T'_1 we arrive at case $A_2)$.

$B, C)$ $\Delta(Ke, Kh) = k^2 > 0$. Then $T = k^{-1}K$ is a symplectic transformation and $\Delta(Kz, Kz') = k^2 \Delta(z, z')$. Let $T_1 = (T T_2)^{-1}$, where T_2 will be chosen later, so that $T_1 K T_2 = kI$.

In case B), $k = 1$ and condition (12.159) is just the positive definiteness of α . We now have the subcases

B₂) If α is nondegenerate, by Lemma 12.12 there is a symplectic transformation T_2 such that

$$\alpha(T_2 z, T_2 z) = N_c (x^2 + y^2),$$

where $N_c > 0$. Also, if $\alpha = 0$, we have a similar formula with $N_c = 0$.

B₁) On the other hand, if α is degenerate of rank one, $\alpha(z, z) = (k_1 x + k_2 y)^2$ for some k_1, k_2 simultaneously not equal to zero, and then for arbitrary $N_c > 0$ there is a symplectic transformation T_2 such that

$$\alpha(T_2 z, T_2 z) = N_c x^2$$

In particular, we can take $N_c = \frac{1}{2}$.

In case C), $k \neq 1$, and condition (12.159) implies that $\alpha/|k^2 - 1|$ is a covariance matrix of a quantum Gaussian state. Hence, there exists a symplectic transformation T_2 such that

$$\alpha(T_2 z, T_2 z) = |k^2 - 1| \left(N_0 + \frac{1}{2} \right) (x^2 + y^2) = \left(N_c + \frac{|k^2 - 1|}{2} \right) (x^2 + y^2),$$

where $N_0 \geq 0$, $N_c = |k^2 - 1| N_0$.

D) $\Delta(K e, K h) = -k^2 < 0$. Then $T = k^{-1} K$ is an antisymplectic transformation, which means that $\Delta(T z, T z') = -\Delta(z, z')$ and $\Delta(K z, K z') = -k^2 \Delta(z, z')$. Similar to cases B,C), we obtain

$$\alpha(T_2 z, T_2 z) = (k^2 + 1) \left(N_0 + \frac{1}{2} \right) (x^2 + y^2) = \left(N_c + \frac{(k^2 + 1)}{2} \right) (x^2 + y^2),$$

with $N_0 \geq 0$, $N_c = (k^2 + 1) N_0$. Letting $T_1 = \Lambda (T T_2)^{-1}$, where $\Lambda [x \ y]^\top = [x \ -y]^\top$, we obtain the first equation in D). Note that T_1 is symplectic, since both T and Λ are antisymplectic.

□

As explained in Section 12.4.1, a Gaussian channel Φ can be dilated to a linear dynamics of an open bosonic system described by the symplectic transformation of the canonical variables q, p and the ancillary canonical variables q_E, p_E, \dots in a Gaussian state S_E . Moreover, this linear dynamics also provides a description of the weak complementary channel Φ_E mapping the initial state of the system q, p into the final state of the environment q'_E, p'_E, \dots . In the case where the state S_E is pure, Φ_E is just the complementary channel $\tilde{\Phi}$, which is determined by Φ up to unitary

equivalence. Note that the environment may have more than one mode (which is reflected by introducing ... in the environment variables).

Let us provide such a description in terms of the input-output relation in each of the cases of Theorem 12.41.

A₁) This is the completely depolarizing channel

$$\begin{aligned} q' &= q_E \\ p' &= p_E \end{aligned}$$

where q_E, p_E describe the environment in the quantum thermal state S_E with mean number of quanta N_0 . Its weak complementary is the ideal channel Id .

A₂) The input-output relation is

$$\begin{aligned} q' &= q + q_E \\ p' &= p_E, \end{aligned} \tag{12.162}$$

where q_E, p_E are again in the quantum thermal state S_E with mean number of quanta N_0 . The equation describes a degenerate classical signal q with additive quantum Gaussian noise. The weak complementary to this channel is described by the transformation

$$\begin{aligned} q'_E &= q \\ p'_E &= p - p_E, \end{aligned} \tag{12.163}$$

where p_E can be regarded as a classical real Gaussian variable with variance $N_0 + \frac{1}{2}$.

B₁) The equation of the channel has the form (12.163), where p_E has variance $\frac{1}{2}$, so that the mode q_E, p_E is in pure (vacuum) state, and the complementary channel is given by (12.162). This is the quantum signal plus degenerate (two-dimensional) classical Gaussian noise.

B₂) Channel with (nondegenerate) additive classical Gaussian noise

$$\begin{aligned} q' &= q + \xi \\ p' &= p + \eta, \end{aligned}$$

where ξ, η are i.i.d. Gaussian random variables with zero mean and variance N_c .

Exercise 12.42. Check that the Stinespring dilation for this channel can be obtained by introducing an environment with two bosonic modes $q_j, p_j; j = 1, 2$ in a pure Gaussian state having zero means, zero covariances, and the variances

$$Dq_1 = Dq_2 = N_c; \quad Dp_1 = Dp_2 = \frac{1}{4N_c},$$

so that $\xi = q_1, \eta = q_2$ and with the following dynamics for the canonical observables:

$$\begin{aligned} q' &= q + q_1, \\ p' &= p + q_2, \\ q'_1 &= q_1, \\ p'_1 &= p_1 - p - q_2/2, \\ q'_2 &= q_2, \\ p'_2 &= p_2 + q + q_1/2. \end{aligned}$$

C) Attenuation/amplification channel with coefficient k and quantum noise with mean number of quanta N_0 . In the attenuation case ($k < 1$), the input-output relation is

$$\begin{aligned} q' &= kq + \sqrt{1-k^2}q_E \\ p' &= kp + \sqrt{1-k^2}p_E, \end{aligned}$$

where q_E, p_E are in the quantum thermal state S_E , with mean number of quanta N_0 . The weak complementary is given by the equations

$$q'_E = \sqrt{1-k^2}q - kq_E \quad (12.164)$$

$$p'_E = \sqrt{1-k^2}p - kp_E, \quad (12.165)$$

and is again an attenuation channel (with coefficient $k' = \sqrt{1-k^2}$).

In the amplification case ($k > 1$), we have

$$\begin{aligned} q' &= kq + \sqrt{k^2-1}q_E \\ p' &= kp - \sqrt{k^2-1}p_E, \end{aligned}$$

with the weak complementary

$$\begin{aligned} q'_E &= \sqrt{k^2-1}q + kq_E \\ p'_E &= -\sqrt{k^2-1}p + kp_E, \end{aligned}$$

see case D).

D) The input-output relation is

$$\begin{aligned} q' &= kq + \sqrt{k^2+1}q_E \\ p' &= -kp + \sqrt{k^2+1}p_E, \end{aligned}$$

which is the same as the weak complementary to the amplification channel with coefficient $k' = \sqrt{k^2+1}$ and quantum noise with mean number of quanta N_0 . This channel describes the attenuation/amplification with phase conjugation, $q, p \rightarrow q, -p$.

Finally, let us show how the c-q channel considered in Section 12.1.4 fits into the general description of quantum Gaussian channels. Since the input is two-dimensional classical, one has to use two bosonic input modes q_1, p_1, q_2, p_2 to describe it quantum-mechanically, so that, e.g. $m_q = q_1, m_p = q_2$. The environment is one mode q, p in the Gaussian state S_0 and hence the output is given by the equations

$$\begin{aligned} q' &= q_1 + q = q + m_q; \\ p' &= q_2 + p = p + m_p, \end{aligned}$$

and the channel parameters are

$$K = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \alpha = \left(N_0 + \frac{1}{2}\right) I_2. \quad (12.166)$$

In this case, the full set of the Heisenberg equations is the same as in case B_2), with the roles of input modes and environmental modes interchanged.

12.6.2 Entanglement-breaking channels

Let us apply theorem 12.35 to the case of one bosonic mode $A = B$, where $\Delta_A(z, z') = \Delta_B(z, z')$ is given by the relation (12.158). As shown above, any one-mode Gaussian channel can be transformed to one of the normal forms.

We only have to find the form $K^\top \Delta_A K$ and check the decomposability (12.135) in each of these cases. We rely upon the simple fact:

Exercise 12.43. Show that

$$\left(N + \frac{1}{2}\right) I_2 \geq \pm \frac{i}{2} \Delta$$

if and only if $N \geq 0$.

With this, we proceed to consideration of each class.

A) $K^\top \Delta K = 0$, hence Φ is a c-q (in fact essentially classical) channel.

B) $K^\top \Delta K = \Delta$, hence the necessary condition of decomposability (12.135) requires $\alpha \geq i \Delta$. This is never satisfied in case B_1) due to the degeneracy of α . Thus, the channel is not entanglement-breaking (in fact it has infinite quantum capacity as shown in [105]). On the other hand, in case B_2) the condition (12.135) is satisfied with $\alpha_B = \alpha_A = \alpha/2$ if and only if $N_c \geq 1$. Hence, Φ is entanglement-breaking in this case.

C) $K^\top \Delta K = k^2 \Delta$. It is clear that in this case the decomposability condition holds if and only if $\alpha \geq \pm \frac{i}{2}(1 + k^2)\Delta$, which is equivalent to $N_c + \frac{|1-k^2|}{2} \geq \frac{(1+k^2)}{2}$ or

$$N_c \geq \min(1, k^2). \quad (12.167)$$

This provides the condition for the entanglement-breaking (which also formally includes the case B_2).

D) $K^\top \Delta K = -k^2 \Delta$. Again, the decomposability condition holds if and only if $\alpha \geq \pm \frac{i}{2}(1 + k^2)\Delta$, which is always the case. Hence, the channel is entanglement-breaking for all $N_c \geq 0$.

Thus, the additivity property (11.33) holds for one-mode Gaussian channels of the form A), D) with arbitrary parameters, and B_2), C) with parameters satisfying (12.167). In general, entanglement-breaking channels have zero quantum capacity, $Q(\Phi) = 0$, see Corollary 10.28. However, in Section 12.6.4 a broader domain of zero quantum capacity will be demonstrated based on degradability analysis.

Exercise 12.44. Use Exercise 12.43 to show that in the case of one mode, condition (12.142) implies (12.135) and hence every PPT channel is entanglement-breaking.

12.6.3 Attenuation/amplification/classical noise channel

From the point of view of applications, cases C) and B_2) are the most interesting ones. The action of the channel Φ can in these cases be described by the single formula

$$\text{Tr } \Phi[S]W(z) = \text{Tr } SW(kz) \cdot \exp \left[-\frac{1}{2} (N_c + |k^2 - 1|/2) (x^2 + y^2) \right], \quad (12.168)$$

where the parameter $N_c \geq 0$ is the power of the environment noise and $k > 0$ is the coefficient of attenuation/amplification in case C). In this case, $N_c = |k^2 - 1|N_0$, where N_0 is the mean photon number of the environment. The case B_2) of an additive classical Gaussian noise channel corresponds to the value $k = 1$. Obviously, the channel is gauge-covariant with respect to the natural complex structure associated with the multiplication by i in the complex plane x, y . Therefore, in what follows, basing ourselves on the results of Section 12.5, we can restrict ourselves to the gauge-invariant states.

Let the input state $S = S(N)$ of the system be the elementary Gaussian with the characteristic function

$$\text{Tr } S(N)W(z) = \exp \left[-\frac{1}{2} \left(N + \frac{1}{2} \right) (x^2 + y^2) \right],$$

the parameter N representing the power (mean number of quanta) of the signal

$$\text{Tr } S(N)a^\dagger a = N. \quad (12.169)$$

In this case, the entropy of $S(N)$ is

$$H(S(N)) = g(N). \quad (12.170)$$

From (12.168) we find that the output state $\Phi[S]$ is the Gaussian state $S(N')$ where

$$N' = k^2 N + N'_0, \quad (12.171)$$

and the total noise

$$N'_0 = \max\{0, (k^2 - 1)\} + N_c \quad (12.172)$$

is the output mean photon number corresponding to the vacuum input state $S(0)$. The first term is nonzero only for $k > 1$, when it represents the amplifier noise. The output entropy is equal to

$$H(\Phi[S(N)]) = g(N'). \quad (12.173)$$

Now, we compute the entropy exchange $H(S(N), \Phi)$. The (pure) input state S_{12} of the extended system $\mathcal{H} \otimes \mathcal{H}_0$ is characterized by the 2×2 -matrix (12.107). According to (12.143), the action of the extended channel $\Phi \otimes \text{Id}$ transforms this matrix into

$$\Delta_{12}^{-1} \alpha'_{12} = \begin{bmatrix} 0 & -(N' + 1/2) & k\sqrt{N^2 + N} & 0 \\ N' + 1/2 & 0 & 0 & k\sqrt{N^2 + N} \\ k\sqrt{N^2 + N} & 0 & 0 & N + 1/2 \\ 0 & k\sqrt{N^2 + N} & -(N + 1/2) & 0 \end{bmatrix},$$

where N' is given by (12.171). From formula (12.97), we deduce $H(S(N), \Phi) = g(|\lambda_1| - \frac{1}{2}) + g(|\lambda_2| - \frac{1}{2})$ where $\pm\lambda_1, \pm\lambda_2$ are the eigenvalues of the matrix in the right-hand side. Solving the characteristic equation, we obtain

$$\lambda_{1,2} = \frac{i}{2} ((N' - N) \pm D), \quad (12.174)$$

where

$$D = \sqrt{(N + N' + 1)^2 - 4k^2 N(N + 1)}.$$

Hence,

$$H(S(N), \Phi) = g\left(\frac{D + N' - N - 1}{2}\right) + g\left(\frac{D - N' + N - 1}{2}\right). \quad (12.175)$$

Let us consider the classical capacities of the channel Φ , under the additive input energy constraint (11.28) corresponding to the operator $F = a^\dagger a$. According to Theorem 12.36 and its corollary, the entanglement-assisted classical capacity

$$C_{ea}(\Phi, F, E) = \max_{S: \text{Tr } S a^\dagger a \leq E} I(S, \Phi)$$

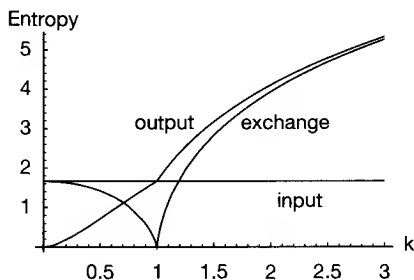


Figure 12.1. Entropies: the output entropy (12.173), the entropy exchange (12.175) for the case $N_c = 0$.

is attained on a gauge-invariant Gaussian state. Since for our channel the condition (11.42) is obviously satisfied, $\text{Tr } S a^\dagger a = E$ for this state, which is thus the elementary Gaussian state $S(E)$. Hence, the maximum is equal to

$$C_{ea}(\Phi, F, E) = I(S(E), \Phi) = H(S(E)) + H(\Phi[S(E)]) - H(S(E), \Phi),$$

where the entropies are given by relations (12.170), (12.173) and (12.175), with N replaced by E .

Considering $C_{ea}(\Phi, F, E)$ as a function of the parameters E, k, N_c , it is interesting to compare it with the quantity $C_\chi(\Phi, F, E)$, which gives a lower bound for the classical capacity $C(\Phi, F, E)$ (and possibly, coincides with it). If the hypothesis of Gaussian optimal ensembles holds true, the optimal ensemble consists of coherent states $S_\zeta = |\zeta\rangle\langle\zeta|$; $\zeta \in \mathbb{C}$, with the Gaussian probability density $p(\zeta) = (\pi E)^{-1} \exp(-|\zeta|^2/E)$, producing the value

$$C_\chi(\Phi, F, E) = g(k^2 E + N'_0) - g(N'_0). \quad (12.176)$$

The ratio

$$G = \frac{C_{ea}(\Phi, F, E)}{C_\chi(\Phi, F, E)} \quad (12.177)$$

then gives at least an upper bound for the *gain* of using entanglement-assisted versus unassisted classical communication.

Exercise 12.45. Show that when the mean number of quanta E in the signal tends to zero, while the total noise $N'_0 > 0$,

$$C_\chi(\Phi, F, E) \sim k^2 E \log \left(\frac{N'_0 + 1}{N'_0} \right),$$

$$C_{ea}(\Phi, F, E) \sim -k^2 E \log E / (N'_0 + 1),$$

so that the entanglement gain G tends to infinity as $-\log E$.

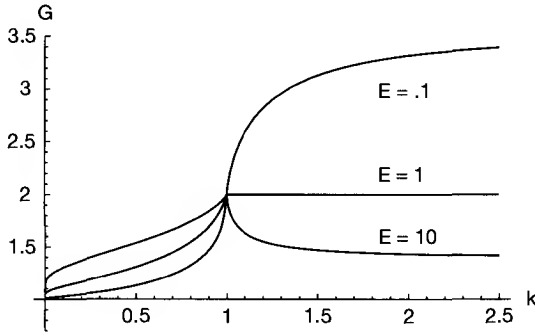


Figure 12.2. The entanglement-assistance gain (12.177) as a function of k for $N_c = 0$. Parameter—the input energy E .

Now, consider the case $N'_0 = 0$, which corresponds to the *ideal attenuator* with zero noise $N_c = 0$ and attenuation coefficient $k < 1$, also called a *pure loss channel* (beam splitter) in quantum optics. According to (12.171) the vacuum state ($N = 0$) is reproduced by this channel, making the minimal output entropy equal to zero and hence additive. In fact, this case falls under the conditions of Proposition 12.40, implying

Corollary 12.46. [67] *For the ideal attenuator, the additivity and the Gaussian optimizer conjectures hold, with*

$$C(\Phi, F, E) = C_\chi(\Phi, F, E) = g(k^2 E). \quad (12.178)$$

Indeed, for this channel $N' = k^2 E$ so that

$$\max_{S: \text{Tr } S a^\dagger a \leq E} H(\Phi[S]) = H(\Phi[S(E)]) = g(k^2 E).$$

□

In the case of the ideal attenuator, we find $D = (1 - k^2)N + 1$, and the entropy exchange (12.175) is equal to $g((1 - k^2)N)$, whence

$$C_{ea}(\Phi, F, E) = g(E) + g(k^2 E) - g((1 - k^2)E).$$

Thus, the gain of entanglement assistance is

$$G = 1 + \frac{g(E) - g((1 - k^2)E)}{g(k^2 E)}.$$

If the signal power $E \rightarrow 0$, using the asymptotic $g(E) \sim -E \log E$, we obtain $C(\Phi, F, E) \sim -k^2 E \log E$, $C_{ea}(\Phi, F, E) \sim -2k^2 E \log E$ so that $G \rightarrow 2$.

Expression (12.178) for the ideal attenuator can be used to provide a tight upper bound for the classical capacity of the channel, defined by formula (12.168) with arbitrary parameters k, N_c . Indeed, denoting this channel by Φ_{k,N_c} one has

$$\Phi_{k,N_c} = \Phi_{\sqrt{N'_0+1},0} \circ \Phi_{k/\sqrt{N'_0+1},0}, \quad (12.179)$$

where N'_0 is given by (12.172), so that $N'_0 + 1 > \max\{1, k^2\}$. In other words, the channel Φ_{k,N_c} can be represented as the ideal attenuator with coefficient $k/\sqrt{N'_0+1}$, followed by the ideal amplifier with coefficient $\sqrt{N'_0+1}$. This is a consequence of (12.168) and the transformation formula (12.171) for the signal power. By using the data-processing inequality

$$C(\Phi_2 \circ \Phi_1, F, E) \leq C(\Phi_1, F, E) \quad (12.180)$$

which is obtained similar to inequality (10.36), but also taking into account the identical input constraints for the channels $\Phi_2 \circ \Phi_1, \Phi_1$, we get

$$C(\Phi_{k,N_c}, F, E) \leq C(\Phi_{k/\sqrt{N'_0+1},0}, F, E).$$

Applying formula (12.178) for the ideal attenuator $\Phi_{k/\sqrt{N'_0+1},0}$ and comparing it with (12.176) results in

$$g(k^2 E + N'_0) - g(N'_0) \leq C(\Phi_{k,N_c}, F, E) \leq g(k^2 E / (N'_0 + 1)). \quad (12.181)$$

The above argument was provided in [137] (in the case of an attenuator, $k < 1$), where it was also shown that the difference between the functions in the left-, and right-hand sides does not exceed $\log e \approx 1.45$ bits.

12.6.4 Estimating the quantum capacity

In this section, the quantum capacity of the one-mode channel will be computed in several cases, based on the (anti-)degradability property (see Section 10.3.3, 12.5.3). We use the fact that composition of one-mode Gaussian channels is again a one-mode Gaussian channel. In particular, the general rule (12.128) implies the following relations.

Denote by $\Phi_{C,k}$ (resp. $\Phi_{D,k}$) a channel of class C (resp. D) with coefficient k and a fixed value of parameter N_c . In this case,

$$\Phi_{C,k_2} \circ \Phi_{C,k_1} = \Phi_{C,k_1 k_2}; \quad \text{if } k_1, k_2 < 1, \quad (12.182)$$

$$\Phi_{D,k_2} \circ \Phi_{C,k_1} = \Phi_{D,k_1 k_2}; \quad \text{if } k_1 > 1. \quad (12.183)$$

Proposition 12.47. *For an attenuator with $k \leq \frac{1}{\sqrt{2}}$, the quantum capacity $Q(\Phi_{C,k}) = 0$. For an attenuation/amplification channel with $k \geq \frac{1}{\sqrt{2}}$ and $N_c = 0$,*

$$Q(\Phi_{C,k}) = Q_G(\Phi_{C,k}) = \log \frac{k^2}{|k^2 - 1|}. \quad (12.184)$$

Here

$$Q_G(\Phi) = \sup_{N \geq 0} I_c(S(N), \Phi),$$

where the supremum of the coherent information $I_c(S, \Phi) = H(\Phi[S]) - H(S, \Phi)$ is taken over all gauge-invariant Gaussian input states $S(N)$.

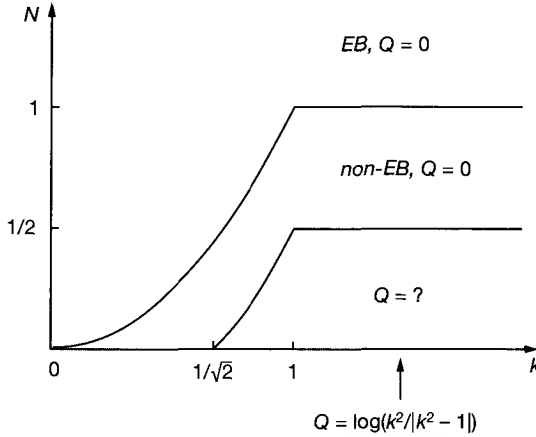


Figure 12.3. Entanglement breaking (EB) and the quantum capacity Q .

Proof. The proof will follow if we show that attenuator with $k < \frac{1}{\sqrt{2}}$ is anti-degradable, while the attenuation/amplification channel with $k \geq \frac{1}{\sqrt{2}}$ and $N_c = 0$ is degradable.

Let $k < \frac{1}{\sqrt{2}}$. In this case, $k_1 = \frac{k}{\sqrt{1-k^2}} < 1$ and, taking into account that the weak complementary of $\Phi_{C,k}$ is

$$\tilde{\Phi}_{C,k}^w = \Phi_{C,\sqrt{1-k^2}}, \quad (12.185)$$

see (12.164), relation (12.182) takes the form

$$\Phi_{C,k} = \Phi_{C,\frac{k}{\sqrt{1-k^2}}} \circ \tilde{\Phi}_{C,k}^w, \quad (12.186)$$

which implies that $\Phi_{C,k}$ is anti-degradable, since the weak complementary can be diluted to the complementary by purifying the environment state. Hence, by Proposition 10.27, its quantum capacity is equal to zero.

In case $N_c = 0$, when the environment state is pure, the complementary channel coincides with the weakly complementary $\tilde{\Phi}_{C,k} = \tilde{\Phi}_{C,k}^w = \Phi_{C,k'}$, where $k' = \sqrt{1-k^2} > \frac{1}{\sqrt{2}}$, and relation (12.186) is the same as

$$\tilde{\Phi}_{C,k'} = \Phi_{C, \frac{k}{\sqrt{1-k^2}}} \circ \Phi_{C,k'},$$

which means that $\Phi_{C,k'}$ is degradable for $\frac{1}{\sqrt{2}} < k' < 1$. Therefore, Proposition 12.38 applies, implying

$$Q(\Phi_{C,k}) = Q_G(\Phi_{C,k}). \quad (12.187)$$

Similarly, in the amplifier case ($k > 1$, $N_c = 0$) relation (12.183) implies

$$\tilde{\Phi}_{C,k} = \Phi_{D, \frac{\sqrt{k^2-1}}{k}} \circ \Phi_{C,k},$$

whence $\Phi_{C,k}$ is degradable and (12.187) continues to hold. It remains for us to compute $Q_G(\Phi_{C,k})$.

In the case where $N_c \geq 0$, the coherent information

$$I_c(S(N), \Phi) = g(N') - g\left(\frac{D + N' - N - 1}{2}\right) - g\left(\frac{D - N' + N - 1}{2}\right) \quad (12.188)$$

is a complicated function of the input power N . We have $I_c(S(0), \Phi) = 0$ and

$$\lim_{N \rightarrow \infty} I_c(S(N), \Phi) = \log k^2 - \log |k^2 - 1| - g(N_c / |k^2 - 1|), \quad k \neq 1.$$

Let us now consider the case where $N_c = 0$. The behavior of the entropies $H(\Phi[S(N)])$, $H(S(N), \Phi)$ as functions of k for $N_c = 0$ is shown in Figure 12.1. Note that, for all N , the coherent information $H(\Phi[S(N)]) - H(S(N), \Phi)$ turns out to be positive for $k^2 > \frac{1}{2}$ and negative otherwise. It tends to $-H(S(N))$ for $k \rightarrow 0$, is equal to $H(S(N))$ for $k = 1$, and quickly tends to zero as $k \rightarrow \infty$.

If $k < 1$ (ideal attenuator), it follows that $N' = k^2 N$, $D = (1 - k^2)N + 1$, and

$$I_c(S(N), \Phi) = g(k^2 N) - g((1 - k^2)N).$$

Exercise 12.48. Show that $I_c(S(N), \Phi)$ is a convex decreasing function of N for $k^2 < \frac{1}{2}$ (correspondingly, a concave increasing function for $k^2 > \frac{1}{2}$), hence

$$\sup_N I_c(S(N), \Phi) = \begin{cases} I_c(S(N), \Phi)|_{N=0} = 0, & k^2 < \frac{1}{2} \\ \lim_{N \rightarrow \infty} I_c(S(N), \Phi) = \log \frac{k^2}{1-k^2}, & k^2 > \frac{1}{2}. \end{cases}$$

In the case where $k > 1$ (ideal amplifier), using (12.171), we have $N' = k^2(N + 1) - 1$, $D = k^2(N + 1) - N$ and

$$I_c(S(N), \Phi) = g(k^2(N + 1) - 1) - g((k^2 - 1)(N + 1)).$$

Exercise 12.49. Show that $I_c(S(N), \Phi)$ is a concave increasing function of N and therefore,

$$\sup_N I_c(S(N), \Phi) = \lim_{N \rightarrow \infty} I_c(S(N), \Phi) = \log \frac{k^2}{k^2 - 1}.$$

Summarizing, we get the general expression (12.184) for the quantum capacity of the attenuator/amplifier with $N_c = 0$. \square

The vast area of zero quantum capacity, containing the domain (12.167) of the entanglement-breaking channels, is given by the following proposition.

Proposition 12.50. Let $k \geq \frac{1}{\sqrt{2}}$. In this case, $Q(\Phi_{C,k}) = 0$ for

$$N_c \geq \frac{1}{2} (k^2 - |k^2 - 1|) = \min\{k^2, 1\} - \frac{1}{2}. \quad (12.189)$$

Proof. Consider the concatenation $\Phi = \Phi_2 \circ \Phi_1$, where

$$\begin{aligned} \Phi_1^*[W(z)] &= W(\sqrt{2}kz) \exp \left[-\frac{(2k^2 - 1)}{2} \left(N_1 + \frac{1}{2} \right) (x^2 + y^2) \right], \\ \Phi_2^*[W(z)] &= W \left(\frac{z}{\sqrt{2}} \right) \exp \left[-\frac{1}{4} \left(N_2 + \frac{1}{2} \right) (x^2 + y^2) \right], \end{aligned}$$

so that, by Proposition 12.47, $Q(\Phi_2) = 0$ and hence $Q(\Phi_2 \circ \Phi_1) = 0$ by (10.35). Then

$$\Phi_1^* \circ \Phi_2^*[W(z)] = W(kz) \exp \left[-\frac{1}{2} \left(\frac{|k^2 - 1|}{2} + N_c \right) (x^2 + y^2) \right],$$

where

$$\frac{|k^2 - 1|}{2} + N_c = \frac{1}{2} \left(N_2 + \frac{1}{2} \right) + \frac{(2k^2 - 1)}{2} \left(N_1 + \frac{1}{2} \right).$$

By varying $N_1, N_2 \geq 0$, one gets all values N_c that satisfy (12.189). \square

In cases A) and D), the channel Φ is anti-degradable [105], and hence also $Q(\Phi) = 0$.

12.7 Notes and references

1. The Spectral Theorem for self-adjoint operators, Stone's Theorem and related questions are considered in detail in the book of Reed and Simon [171].

The quantum harmonic oscillator, first studied by Dirac [53], underlies many mathematical models in quantum optics and quantum electronics. In particular, useful symbolic calculus leading to algebraic identities following from CCR, such as (12.15), was developed in the book of Louisell [151]. An analysis based on coherent states was developed in the classical work of Glauber [69]. Relation (12.22) is the famous P -representation, widely used in quantum optics. The representation of the radiation field as an ensemble of quantum oscillators, introduced by Dirac, is discussed in the books of Klauder and Sudarshan [136] and Helstrom [86], in a form suitable for applications in quantum optics.

By using overcompleteness of the system of coherent vectors (Exercise 12.6)), one can define the unsharp observable with values in $\mathbb{C} \equiv \mathbb{R}^2$

$$M(B) = \frac{1}{\pi} \int_B |\zeta\rangle\langle\zeta| d^2\zeta, \quad (12.190)$$

which plays a key role in the description of an “approximate joint measurement” of observables q and p . A detailed discussion and further references can be found in the books of Davies [48] and Holevo [107].

The formula for the capacity of a c-q Gaussian channel was conjectured by Gordon [71]. In our proof, we follow [98], where one can also find a consideration of the multimode Gaussian c-q channel, including a realistic stochastic process model of the classical signal on the background of colored quantum noise. The capacity of the channel with squeezed Gaussian noise was computed by Holevo, Sohma, and Hirota in [112]. Much more detailed information concerning the rate of convergence of the error probability for pure state Gaussian channels can be obtained by modifying the estimates from Section 5.7 to channels with infinite alphabets and constrained input [113].

2. The usefulness of the symmetric description (12.47) of CCR was stressed by Segal [181]. For the proof of the following theorem, see, e.g. [107].

3. Stone-von Neumann Uniqueness Theorem *Let V_x, U_y ; $x, y \in \mathbb{R}^s$ be two strongly continuous families of unitary operators in a separable Hilbert space \mathcal{H} satisfying the Weyl CCR (12.44). In this case, V_x, U_y are unitarily equivalent to a direct sum of at most a countable number of copies of the Schrödinger representation arising from (12.43). In particular, any irreducible representation is unitarily equivalent to the Schrödinger representation.*

The paper of Williamson [222] contains general results concerning normal forms of not necessarily positive quadratic Hamiltonians. For the proof of Lemma 12.12, see also Chapter V of the book [107].

Many physical applications of the linear dynamics of systems with quadratic Hamiltonians are considered in the book of Malkin and Man'ko [155]. A detailed review of the group of real symplectic transformations, with applications to quantum optics and mechanics, is given by Arvind, Dutta, Mukunda, and Simon [10].

4. In our exposition of quantum Gaussian states, we follow Chapter V of the book [107], where the proof of Exercises 12.15, 12.16, 12.19, and of sufficiency of the condition (12.76) can be found. In [107], one can also find an accurate treatment of the expectations of unbounded operators and moments of a quantum state. The approach to quantum Gaussian states via characteristic functions is based on apparent analogies with probability theory and is perhaps the most direct and transparent. An alternative way used in physical papers is via the Wigner distribution function, see, e.g. [195].

The formulas for the entropy of a general quantum state were obtained in [112]. The maximum entropy characterization of the Gaussian states (Lemma 12.25) is a particular case of the maximum entropy principle in statistical mechanics, see e.g. [151], Section 6.6. A similar property of the conditional quantum entropy (Lemma 12.26) was noticed by Eisert and Wolf [55].

The separability criterion for Gaussian states was obtained in the paper of Werner and Wolf [217]. This paper also contains important results concerning Gaussian PPT (positive partial transpose) states, in particular, a proof that in $1 \times N$ bosonic modes, the PPT is sufficient for separability of Gaussians and the construction of a family of Gaussian PPT nonseparable (bound entangled) states in a 2×2 system. Purification of the Gaussian state was computed in [90].

5. The general description of bosonic linear and Gaussian channels was given in Holevo and Werner [114] (see also the survey [55]), following [92]. By making a canonical decomposition for a general bosonic Gaussian channel studied in this section, it is possible to show that an arbitrary quantum optical communication circuit can be seen as composed of basic building blocks, comprising linear multiport interferometers and a few basic one-mode “non-linear” (mathematically – non-gauge covariant) operations, such as parametric down-converters or squeezers [31].

On the other hand, the Gaussian channels are a special case of quasi-free maps of the CCR-algebra considered by Demoen, Vanheuverzwijn, and Verbeure [50], who had shown that nonnegative definiteness of the matrices (12.125) is necessary and sufficient for complete positivity of the maps Φ . In [50], the problem of unitary dilation of a quasi-free map was considered under a rather special restriction on the operator K . The proof of Theorem 12.30 contains the construction of the unitary dilation for

an arbitrary Gaussian channel, see Caruso, Eisert, Giovannetti, and Holevo [37] for more detail.

Gaussian observables are essentially Gaussian q-c channels. They are considered in Chapter VI of the book [107]. In particular, it is shown that the Naimark dilation for the unsharp observable (12.190) is given by the joint spectral measure of commuting operators $q + q_C$, $p - p_C$, where the mode C is in the vacuum state.

Gaussian entanglement-breaking channels were studied in the work of Holevo [106].

6. The principle “Gaussian channels have Gaussian optimizers” is well known in classical information theory. The hypothesis of optimal Gaussian ensembles extrapolates this principle to quantum Gaussian states and channels, but only partial results are known in this direction. There are indications that the confirmation of this hypothesis should depend on the solution of the additivity problem for quantum Gaussian channels. In the paper of Wolf, Giedke, and Cirac [226], it is shown that if the additivity holds, the average state \bar{S}_π of the optimal ensemble should be Gaussian. A closely related conjecture of the Gaussian minimizer for the entanglement of formation has a positive solution for symmetric Gaussian states in 1×1 composite bosonic system, as shown by Wolf et al. [227].

Theorem 12.36, proved in Holevo and Werner [114], is one instance of this principle for the mutual quantum information. Proposition 12.38, which expresses this principle for the quantum capacity of degradable Gaussian channels, is based on the observation of Wolf, Pérez-García and Giedke [228].

Proposition 12.40 is a multimode generalization of a result of Giovannetti, Guha, Lloyd, Maccone, Shapiro, and Yuen [67] (Corollary 12.46) for a pure loss channel.

7. The classification of one-mode Gaussian channels was obtained by Holevo [106], see also [36]. A detailed study of their structure, based on explicitly found Kraus decompositions, is given in the paper of Ivan, Sabapathy, and Simon [122].

The capacities of the attenuation/amplification channel were considered by Holevo and Werner [114]. The upper bound (12.181) for the classical capacity of the attenuation channel was proposed by König and Smith [137], who also derived still better bounds based on a quantum generalization of the entropy power inequality in information theory.

The fact that the quantum capacity is given by the Gaussian expression obtained in [114] was observed by Wolf, Pérez-García and Giedke [228]. The regions of zero quantum capacity were described in Caruso, Giovannetti, and Holevo [36]. Solutions to the Exercises from Section 12.6.4 and additional information on the channel with additive classical Gaussian noise are given in the paper [105]. Smith, Smolin, and Yard [199] demonstrated superactivation of the quantum capacity for the pair of zero-capacity Gaussian channels: a one-mode attenuator with $k^2 = 1/2$ and an explicitly constructed two-mode PPT channel.

Numerous physical applications of continuous variable (in particular bosonic Gaussian) information processing systems, including experimental realizations, are discussed in the surveys [32] and [214].

Bibliography

- [1] A. Abeyesinghe, I. Devetak, P. Hayden, A. Winter, *The mother of all protocols: restructuring quantum information's family tree*, arXiv:quant-ph/0606225 (2006).
- [2] C. Adami, N.J. Cerf, *Capacity of noisy quantum channels*, Phys. Rev. A **56** (1997), pp. 3470–3485.
- [3] R. Ahlswede, I. Csiszár, *Common randomness in information theory and cryptography – Part II: CR capacity*, IEEE Trans. Inform. Theory **44** (1998), pp. 225–240.
- [4] R. Ahlswede, A. Winter, *Strong converse for identification via quantum channels*, IEEE Trans. Inf. Theory **48** (2002), pp. 569–579.
- [5] G. Alber, T. Beth, M. Horodecki, P. Horodecki, R. Horodecki, M. Rötteler, H. Weinfurter, R.F. Werner, *Quantum Information. An Introduction to Basic Theoretical Concepts and Experiments*, Springer, Berlin, 2001.
- [6] R. Alicki, *Isotropic quantum spin channels and additivity questions*, arXiv:quant-ph/0402080 (2004).
- [7] H. Araki, *On a characterization of the state space in quantum mechanics*, Commun. Math. Phys. **75** (1980), pp. 1–24.
- [8] P.K. Aravind, *Quantum mysteries revisited again*, Am. J. Phys. **72** (2004), pp. 1303–1307.
- [9] W.B. Arveson, *Subalgebras of C^* -algebras*, Acta Math. **123** (1969), pp. 141–224.
- [10] Arvind, B. Dutta, N. Mukunda, R. Simon, *The real symplectic groups in quantum mechanics and optics*, Pramana **45** (1995), pp. 471.
- [11] G. Aubrun, S. Szarek, E. Werner, *Hastings' additivity counterexample via Dvoretzky's theorem*, arXiv:1003.4925 [quant-ph] (2010).
- [12] K.M.R. Audenaert, *A sharp Fannes-type inequality for the von Neumann entropy*, arXiv:quant-ph/0610146 (2006).
- [13] K.M.R. Audenaert, S.L. Braunstein, *On strong superadditivity of the entanglement of formation*, Commun. Math. Phys. **246** (2004), pp. 443–452.
- [14] P.A. Bakut, S.S. Shchurov, *Optimal detection of quantum signal*, Probl. Inform. Transmission **4** (1968), pp. 61–65.
- [15] H. Barnum, E. Knill, M.A. Nielsen, *On quantum fidelities and channel capacities*, IEEE Trans. Inform. Theory **46** (1998), pp. 1317–1329; arXiv:quant-ph/9809010.
- [16] H. Barnum, M.A. Nielsen, B. Schumacher, *Information transmission through a noisy quantum channel*, Phys. Rev. A **57** (1998), pp. 4153–4175.

- [17] V. P. Belavkin, *Optimal multiple quantum statistical hypotheses testing*, J. Stoch. **1** (1975), pp. 315–345.
- [18] J. S. Bell, *On the Einstein–Podolsky–Rosen paradox*, Physics **1** (1964), pp. 195–200.
- [19] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, *Teleporting an unknown quantum state via dual classical and EPR channels*, Phys. Rev. Lett. **70** (1993), pp. 1855–1859.
- [20] C. H. Bennett, I. Devetak, P. W. Shor, J. A. Smolin, *Inequalities and separations among assisted capacities of quantum channels*, arXiv:quant-ph/0406086 (2004).
- [21] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, W. K. Wootters, *Mixed state entanglement and quantum error correction*, arXiv:quant-ph/9604024 (1996).
- [22] C. H. Bennett, C. A. Fuchs, J. A. Smolin, *Entanglement-enhanced classical communication on a noisy quantum channel*, in *Quantum Communication, Computing and Measurement*, Proc. QCM96, ed. by O. Hirota, A. S. Holevo, C. M. Caves, Plenum, New York, 1997, pp. 79–88; arXiv:quant-ph/9611006.
- [23] C. H. Bennett, P. W. Shor, *Quantum information theory*, IEEE Trans. Inform. Theory **44** (1998), pp. 2724–2742.
- [24] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal, *Entanglement-assisted classical capacity of noisy quantum channel*, Phys. Rev. Lett. **83** (1999), 3081; arXiv:quant-ph/9904023.
- [25] C. H. Bennett, P. W. Shor, J. A. Smolin, A. V. Thapliyal, *Entanglement-assisted capacity and the reverse Shannon theorem*, IEEE Trans. Inform. Theory **48** (2002), pp. 2637–2655; arXiv:quant-ph/0106052.
- [26] C. H. Bennett, S. J. Wiesner, *Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states*, Phys. Rev. Lett. **69** (1992), pp. 2881–2884.
- [27] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.
- [28] I. Bjelaković, R. Siegmund-Schultze, *A new proof of the monotonicity of quantum relative entropy for finite quantum systems*, arXiv:quant-ph/0307170 (2003).
- [29] G. Bowen, *Quantum Feedback Channels*, IEEE Trans. Inform. Theory **50** (2004), pp. 2429–2433.
- [30] F. G. L. S. Brandão, M. Horodecki, *On Hastings’ counterexample to the minimum output entropy additivity conjectures*, arXiv:0907.3210 [quant-ph] (2009).
- [31] S. L. Braunstein, *Squeezing as an irreducible resource*, arXiv:quant-ph/9904002 (1999).
- [32] S. L. Braunstein, P. van Loock, *Quantum information with continuous variables*, Rev. Mod. Phys. **77** (2005), pp. 514–577.
- [33] M. V. Burnashev, A. S. Holevo, *On reliability function of quantum communication channel*, Probl. Inform. Transm. **34** (1998) pp. 97–107.
- [34] N. Cai, A. Winter, R. W. Yeung, *Quantum privacy and quantum wiretap channels*, Probl. Inform. Transmission **40** (2004), pp. 318–336.

- [35] R. Calderbank, P. W. Shor, *Good quantum error-correcting codes exist*, Phys. Rev. A **54** (1996), pp. 1098–1105.
- [36] F. Caruso, V. Giovannetti, A. S. Holevo, *One-mode Bosonic Gaussian channels: a full weak-degradability classification*, New J. Phys. **8** (2006), 310; arXiv:quant-ph/0609013.
- [37] F. Caruso, J. Eisert, V. Giovannetti, A. S. Holevo, *Multi-mode bosonic Gaussian channels*, New J. Phys. **10** (2008), 083030 (33 pp); arXiv:0804.0511 [quant-ph].
- [38] C. M. Caves, P. B. Drummond, *Quantum limits of bosonic communication rates*, Rev. Mod. Phys. **66** (1994), pp. 481–538.
- [39] N. N. Chentsov, *Selected works. Mathematics. Part IV. Geometrostatistics and non-commutative probability*, Fizmatlit, Moscow, 2001. (in Russian)
- [40] G. Chiribella, G. M. D'Ariano, P. Perinotti, *Informational derivation of quantum theory*, Phys. Rev. A **84** (2011), 012311; arXiv:1011.6451 [quant-ph].
- [41] M.-D. Choi, *Completely positive maps on complex matrices*, Linear Alg. and Its Appl. **10** (1975), pp. 285–290.
- [42] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd edition, John Wiley & Sons, New York, 1996.
- [43] T. S. Cubitt, M.-B. Ruskai, G. Smith, *The structure of degradable quantum channels*, arXiv:0802.1360 [quant-ph] (2008).
- [44] I. Csiszár, J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Akadémiai Kiadó, Budapest, 1981.
- [45] M. Dalai, *Sphere packing and zero-rate bounds to the reliability of classical-quantum channel*, arXiv:1201.5411 [cs.IT] (2012).
- [46] M. Dall'Arno, G. M. D'Ariano, M. F. Sacchi, *Informational power of quantum measurements*, arXiv:1103.1972 [quant-ph] (2011).
- [47] E. B. Davies, *Information and quantum measurement*, IEEE Trans. Inform. Theory **24** (1978), pp. 596–599.
- [48] E. B. Davies, *Quantum Theory of Open Systems*, Academic Press, London, 1976.
- [49] G. F. Dell'Antonio, *On the limits of sequences of normal states*, Commun. Pure Appl. Math. **20** (1967), pp. 413–430.
- [50] B. Demoen, P. Vanheuverzwijn, A. Verbeure, *Completely positive quasi-free maps on the CCR algebra*, Rep. Math. Phys. **15** (1979), pp. 27–39.
- [51] I. Devetak, *The private classical information capacity and quantum information capacity of a quantum channel*, arXiv:quant-ph/0304127 (2003).
- [52] I. Devetak, P. Shor, *The capacity of a quantum channel for simultaneous transition of classical and quantum information*, arXiv:quant-ph/0311131 (2003).
- [53] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th edition, Oxford Univ. Press, Oxford, 1958, (1930).

- [54] D. Di Vincenzo, P. W. Shor, J. Smolin, *Quantum channel capacities of very noisy channels*, Phys. Rev. A **57** (1998), pp. 830–839.
- [55] J. Eisert, M. M. Wolf, *Gaussian quantum channels*, arXiv:quant-ph/0505151 (2005).
- [56] E. G. Effros, *A matrix convexity approach to some celebrated quantum inequalities*, Proc. Natl. Acad. Sci. USA **106** (2009), pp. 1006–1008; arXiv:math-ph/0802.1234.
- [57] L. D. Faddeev, O. A. Yakubovskii, *Lectures on Quantum Mechanics for Mathematics Students* (Student Mathematical Library), AMS, 2009.
- [58] M. Fannes, *A continuity property of quantum entropy for spin lattice systems*, Commun. Math. Phys. **31** (1973), pp. 291–294.
- [59] R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman Lectures on Physics*, vol. 3, Addison–Wesley Publishing Company, Inc., Reading, Massachusetts, Paolo Alto, London, 1965.
- [60] G. D. Forney, Jr., S. M. Thesis, MIT, 1963.
- [61] *Foundations of Quantum Mechanics and Ordered Linear Spaces*. Eds. A. Hartkammer, H. Neumann, Lect. Notes Phys. **29**, Springer-Verlag, New York–Heidelberg–Berlin, 1974.
- [62] C. A. Fuchs, *Distinguishability and accessible information in quantum theory* arXiv:quant-ph/9601020 (1996).
- [63] C. A. Fuchs, *Quantum mechanics as quantum information (and only a little more)*, arXiv:quant-ph/0205039 (2002).
- [64] M. Fukuda, C. King, D. Moser, *Comments on Hastings’ additivity counterexamples*, arXiv:0905.3697 (2009).
- [65] M. Fukuda, M. M. Wolf, *Simplifying additivity problems using direct sum constructions*, J. Math. Phys. **48** (2007), 072101.
- [66] R. Gallager, *Information Theory and Reliable Communications*, Wiley, New York, 1968.
- [67] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, H. P. Yuen, *Classical capacity of the lossy bosonic channel: the exact solution*, Phys. Rev. Lett. **92** (2004), 027902; arXiv:quant-ph/0308012.
- [68] I. M. Glazman, Ju. I. Ljubich, *Finite-Dimensional Linear Analysis: A Systematic Presentation in Problem Form*, MIT Press, Cambridge, Massachusetts, 1974.
- [69] R. J. Glauber, *The quantum theory of optical coherence*, Phys. Rev. **130** (1963), pp. 2529–2539.
- [70] J. P. Gordon, *Quantum effect and communications systems*, Proc. IRE **50** (1962), pp. 1898–1908.
- [71] J. P. Gordon, *Noise at optical frequencies; information theory*, In: *Quantum Electronics and Coherent Light*, Proc. Int. School Phys. “Enrico Fermi”, Course XXXI, ed. P. A. Miles, Academic Press, New York, 1964, pp. 156–181.
- [72] D. Gottesman, *Stabilizer codes and quantum error correction*, Ph. D. Thesis, Caltech 1997; arXiv:quant-ph/9705052.

- [73] A. Grudka, M. Horodecky, L. Pankowski, *Constructive counterexamples to additivity of minimum output Rényi entropy of quantum channels for all $p > 2$* , arXiv:0911.2515 [quant-ph] (2009).
- [74] M. J. W. Hall, *Quantum information and correlation bounds*, Phys. Rev. A **55** (1997), pp. 1050–2947.
- [75] L. Hardy, *Reformulating and reconstructing quantum theory*, arXiv:1104.2066 [quant-ph] (2011).
- [76] M. B. Hastings, *A counterexample to additivity of minimum output entropy*, Nature Phys. **5** (2009), 255; arXiv:quant-ph/0809.3972.
- [77] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. Wootters: *Classical information capacity of a quantum channel*, Phys. Rev. A **54** (1996), pp. 1869–1876.
- [78] M. Hayashi, *Quantum Information: an Introduction*, Springer, Berlin, 2006.
- [79] M. Hayashi, H. Nagaoka, *General formulas for capacity of classical-quantum channels*, IEEE Trans. Inform. Theory **49** (2003), pp. 1753–1768; arXiv:quant-ph/0206186.
- [80] P. Hayden, *The maximal p -norm multiplicativity conjecture is false*, arXiv:quant-ph/0707.3291 (2007).
- [81] P. Hayden, *Entanglement in random subspaces*, Proc. of QCMC04. AIP conference proceedings, vol. **734**, pp. 226–229, 2004; arXiv:quant-ph/0409157.
- [82] P. Hayden, D. Leung, P. W. Shor, A. Winter, *Randomizing quantum states: constructions and applications*, Communications in Mathematical Physics **250** (2004), pp. 371–391.
- [83] P. Hayden, A. Winter, *Counterexamples to the maximal p -norm multiplicativity conjecture for all $p > 1$* , Comm. Math. Phys. **284** (2008), pp. 263–280; arXiv:0807.4753 [quant-ph].
- [84] P. Hayden, P. W. Shor, A. Winter, *Random quantum codes from Gaussian ensembles and an uncertainty relation*, Open Syst. Inf. Dyn. **15** (2008), pp. 71–89; arXiv:quant-ph/0712.0975.
- [85] C. W. Helstrom, *Detection theory and quantum mechanics*, Inform. and Control **10** (1967), pp. 254–291.
- [86] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976.
- [87] U. Herzog, *Optimum state discrimination with a fixed rate of inconclusive results: Analytical solutions and relation to state discrimination with a fixed error rate*, arXiv:1206.4412 [quant-ph].
- [88] T. Hiroshima, *Additivity and multiplicativity properties of some Gaussian channels for Gaussian inputs*, Phys. Rev. A **73** (2006), 012330; arXiv:quant-ph/0511006.
- [89] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), pp. 13–30.

- [90] A. S. Holevo, *Generalized free states of the C^* -algebra of the CCR*, Theor. Math. Phys. **6** (1971), N1, pp. 1–12.
- [91] A. S. Holevo, *Statistical problems in quantum physics*, 2nd Japan–USSR Symp. on Probability Theory **1**, Kyoto (1972), pp. 22–40.
- [92] A. S. Holevo, *Towards a mathematical theory of quantum communication channels*, Probl. Inform. Transm. **8** (1972), pp. 63–71.
- [93] A. S. Holevo, *On quasiequivalence of locally normal states*, Theor. Math. Phys., **13** (1972), pp. 184–199.
- [94] A. S. Holevo, *Information-theoretic aspects of quantum measurement*, Probl. Inform. Transm. **9** (1973), pp. 31–42.
- [95] A. S. Holevo, *Some estimates for the information content transmitted by a quantum communication channel*, Probl. Inform. Transm. **9** (1973), pp. 3–11.
- [96] A. S. Holevo, *Statistical definition of observable and the structure of statistical models*, Rep. Math. Phys. **22** (1985), pp. 385–407.
- [97] A. S. Holevo, *The capacity of quantum communication channel with general signal states*, IEEE Trans. Inform. Theory **44** (1998), pp. 269–272; arXiv:quant-ph/9611023 (1996).
- [98] A. S. Holevo, *Quantum coding theorems*, Russian Math. Surveys **53** (1998), pp. 1295–1331; arXiv: quant-ph/9809023.
- [99] A. S. Holevo, *Statistical Structure of Quantum Theory*, Lect. Notes Phys. **m67** Springer, Berlin, 2001.
- [100] A. S. Holevo, *An Introduction to Quantum Information Theory*, MCCME (Moscow Independent University), Moscow, 2002. (in Russian)
- [101] A. S. Holevo, *On entanglement-assisted classical capacity*, J. Math. Phys. **43**, (2002), pp. 4326–4333; arXiv:quant-ph/0106075.
- [102] A. S. Holevo, *Classical capacities of quantum channels with constrained inputs*, Probab. Theory and Appl. **48** (2003), pp. 359–374; arXiv:quant-ph/0211170.
- [103] A. S. Holevo, *On complementary channels and the additivity problem*, Probab. Theory and Appl. **51** (2006), pp. 133–134; arXiv:quant-ph/0509101.
- [104] A. S. Holevo, *The additivity problem in quantum information theory*, Proc. ICM, Madrid, Spain, 2006, pp. 999–1018.
- [105] A. S. Holevo, *One-mode quantum Gaussian channels*, Probl. Inform. Transmission **43** (2007), pp. 1–11; arXiv:quant-ph/0607051.
- [106] A. S. Holevo, *Entanglement-breaking channels in infinite dimensions*, Probl. Inform. Transmission **44** (2008), pp. 3–18; arXiv:quant-ph/0802.0235.
- [107] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, 2nd edition, Edizioni della Normale, Pisa, 2011.
- [108] A. S. Holevo, *Information capacity of quantum observable*, Probl. Inform. Transmission, arXiv:1103.2615 [quant-ph].

- [109] A. S. Holevo, M. E. Shirokov, *On Shor's channel extension and constrained channels*, Commun. Math. Phys. **249** (2004), pp. 417–430; arXiv:quant-ph/0306196.
- [110] A. S. Holevo, M. E. Shirokov, *Continuous ensembles and the χ -capacity of infinite dimensional channels*, Probab. Theory and Appl. **50** (2005), pp. 98–114.
- [111] A. S. Holevo, M. E. Shirokov, R. F. Werner, *On the notion of entanglement in infinite dimensional spaces*, Russian Math. Surveys **60** (2005), pp. 153–154.
- [112] A. S. Holevo, M. Sohma, O. Hirota, *Capacity of quantum Gaussian channels*, Phys. Rev. A **59** (1999), pp. 1820–1828.
- [113] A. S. Holevo, M. Sohma, O. Hirota, *Error exponents for quantum channels with constrained inputs*, Rep. Math. Phys. **46** (2000), pp. 343–358.
- [114] A. S. Holevo, R. F. Werner, *Evaluating capacities of Bosonic Gaussian channels*, Phys. Rev. A **63** (2001), 032312; arXiv:quant-ph/9912067 (1999).
- [115] M. Horodecki, P. Horodecki, R. Horodecki, *Separability of mixed states: necessary and sufficient conditions*, Phys. Lett A **223** (1996), pp. 1–8.
- [116] P. Horodecki, *Separability criterion and inseparable mixed states with positive partial transposition*, arXiv:quant-ph/9703004 (1997).
- [117] M. Horodecki, P. Horodecki, R. Horodecki, *Unified approach to quantum capacities: towards quantum noisy coding theorem*, Phys. Lett. **85** (2000), pp. 433–436.
- [118] P. Horodecki, M. Horodecki, R. Horodecki, *Binding entanglement channels*, J. Mod. Opt. **47** (2000), pp. 347–354.
- [119] R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, *Quantum entanglement*, Rev. Mod. Phys. **81** (2009), pp. 865–942.
- [120] M. Horodecki, J. Oppenheim, A. Winter, *Quantum state merging and negative information*, Comm. Math. Phys. **269** (2006), pp. 107; arXiv:quant-ph/0512247
- [121] M. Horodecki, P. W. Shor, M. B. Ruskai, *General entanglement breaking channels*, Rev. Math. Phys. **15** (2003), pp. 629–641; arXiv:quant-ph/0302031.
- [122] J. S. Ivan, K. Sabapathy, R. Simon, *Operator-sum representation for Bosonic Gaussian channels*, Phys. Rev. A **84** (2011), 042311.
- [123] A. Jamiolkowski, *Linear transformations which preserve trace and positive semidefiniteness of operators*, Rep. Math. Phys. **3** (1972), 275–278.
- [124] R. Jozsa, D. Robb, W. K. Wootters, *Lower bound for accessible information in quantum mechanics*, Phys. Rev. A **9** (1994), pp. 668–677.
- [125] R. Jozsa, B. Schumacher: *A new proof of the quantum noiseless coding theorem*, J. Modern Optics **41** (1994), pp. 2343–2349.
- [126] K. Kato, M. Osaki, O. Hirota, *Derivation of classical capacity of quantum channels for discrete information sources*, arXiv:quant-ph/9811085.
- [127] K. Kato, M. Osaki, T. Suzuki, M. Ban, O. Hirota, *Upper bound of the accessible information and lower bound of the Bayes cost in quantum signal detection processes*, Phys. Rev. A **54** (1996), 2718–2727.

- [128] M. Keyl, *Fundamentals of quantum information theory* Phys. Rep. **369** (2002), pp. 431–548, arXiv:quant-ph/0202122.
- [129] C. King, *Additivity for unital qubit channels*, J. Math. Phys. **43** (2002), pp. 4641–4643; arXiv:quant-ph/0103156.
- [130] C. King, *The capacity of quantum depolarizing channel*, IEEE Trans. Inf. Th. **49** (2003), pp. 221–229; arXiv:quant-ph/0204172.
- [131] C. King, *Maximal p -norms of entanglement breaking channels*, arXiv:quant-ph/0212057.
- [132] C. King, K. Matsumoto, M. Natanson, M. B. Ruskai, *Properties of conjugate channels with applications to additivity and multiplicativity*, Markov Process and Related Fields **13** (2007), pp. 391–423; arXiv:quant-ph/0509126.
- [133] E. Knill, R. Laflamme, *Theory of quantum error-correcting codes*, Phys. Rev. A **55** (1997), pp. 900–911.
- [134] A. Yu. Kitaev, A. H. Shen', M. N. Vyalyi, *Classical and Quantum Computation* (Graduate Studies in Mathematics), AMS, 2002.
- [135] A. Yu. Kitaev, *Quantum computations: algorithms and error-correction*, Russian Math. Surveys **52** (1997), pp. 1191–1249.
- [136] J. R. Klauder, E. C. G. Sudarshan, *Fundamentals of Quantum Optics*, Benjamin, New York, 1968.
- [137] R. König, G. Smith, *Limits on classical communication from quantum entropy power inequalities*, arXiv:1205.3407 [quant-ph].
- [138] A. I. Kostrikin, Yu. I. Manin, *Linear Algebra and Geometry*, Gordon and Breach Scientific Publications, 1989.
- [139] K. Kraus, *States, Effects and Operations*, Springer Lecture Notes in Physics **190** 1983.
- [140] D. Kretschmann, R. Werner, *Tema con variazioni: quantum channel capacities*, New J. Phys. **6** (2004), 26; arXiv:quant-ph/0311037.
- [141] D. Kretschmann, R. Werner, *Quantum channels with memory* Phys. Rev. A **72** (2005), 062323; arXiv:quant-ph/0502106.
- [142] D. S. Lebedev, L. B. Levitin, *The maximal amount of information transmissible by an electromagnetic field*, Doklady AN SSSR **149** (1963), pp. 1299–1303.
- [143] A. Lesniewski, M. B. Ruskai, *Monotone Riemannian metrics and relative entropy on noncommutative probability spaces*, J. Math. Phys. **40** (1999), pp. 5702–5724.
- [144] L. B. Levitin, *Optimal quantum measurement for two pure and mixed states*, In: Quantum Communications and Measurement, Proc. QCM94, ed. by V. P. Belavkin, O. Hirota, R. L. Hudson, Plenum, New York 1995, pp. 439–448.
- [145] E. H. Lieb, M. B. Ruskai, *Proof of the strong subadditivity of quantum mechanical entropy*, J. Math. Phys., **14** (1973), pp. 1938–1941.
- [146] G. Lindblad, *Entropy, information and quantum measurements*, Commun. Math. Phys. **33** (1973), pp. 305–322.

- [147] G. Lindblad, *Expectations and entropy inequalities for finite quantum systems*, Commun. Math. Phys. **39** (1974), pp. 111–119.
- [148] G. Lindblad, *Completely positive maps and entropy inequalities*, Commun. Math. Phys. **40** (1975), pp. 147–151.
- [149] G. Lindblad, *Quantum entropy and quantum measurements*, Lect. Notes Phys., **378**, Quantum Aspects of Optical Communication, Ed. by C. Benjaballah, O. Hirota, S. Reynaud, 1991, pp. 71–80.
- [150] S. Lloyd, *Capacity of noisy quantum channel*, Phys. Rev. A **55** (1997), pp. 1613–1622.
- [151] W. Louisell, *Radiation and Noise in Quantum Electronics*, McGraw–Hill, New York, 1964.
- [152] G. Ludwig, *Foundations of Quantum Mechanics I*, Springer, Berlin–Heidelberg–New York, 1983.
- [153] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics*, Benjamin, New York, 1963.
- [154] G. G. Magaril–Ilyayev, V. M. Tikhomirov, *Convex Analysis: Theory and Applications* (Translations of Mathematical Monographs **222**), AMS, 2003.
- [155] I. A. Malkin, V. I. Man'ko, *Dynamical Symmetries and Coherent States of Quantum Systems*, Nauka, Moscow, 1979, (in Russian).
- [156] Ll. Masanes, M. P. Müller, *A derivation of quantum theory from physical requirements*, New J. Phys. **13** (2011), 063001.
- [157] K. Matsumoto, T. Shimono, A. Winter, *Remarks on the additivity of the Holevo channel capacity and of the entanglement of formation*, Commun. Math. Phys. **246** (2004) pp. 427–442; arXiv:quant-ph/0206148 (2002).
- [158] M. A. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
- [159] M. A. Naimark, *Spectral functions of a symmetric operator*, Izv. Acad. nauk SSSR Ser. Mat. **4** (1940), pp. 277–318. (in Russian)
- [160] T. Ogawa, H. Nagaoka, *Strong converse and Stein's lemma in quantum hypothesis testing*, IEEE Trans. Inf. Theory, **46** (2000), pp. 2428–2433.
- [161] M. Ohya, D. Petz, *Quantum Entropy and Its Use*, Texts and Monographs in Physics, Springer, New York, 1993.
- [162] M. Ozawa, *On information gain by quantum measurement of continuous observable*, J. Math. Phys. **27** (1986), pp. 759–763.
- [163] M. Ozawa, *Quantum measuring process of continuous observables*, J. Math. Phys. **18** (1987), pp. 412–421.
- [164] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York and London, 1967.
- [165] A. Peres, Phys. Lett. A **128** (1988), pp. 19.
- [166] A. Peres, Phys. Rev. Lett. **76** (1997), pp. 1413.

- [167] D. Petz, *Quasi-entropies for finite quantum systems*, Rep. Math. Phys. **21** (1986), pp. 57–65.
- [168] D. Petz, *Quantum Information Theory and Quantum Statistics*, Springer, Berlin, 2008.
- [169] R. T. Powers, E. Störmer, *Free states on the canonical anticommutation relations*, Commun. Math. Phys. **16** (1970), pp. 1–33.
- [170] E. M. Rains, *Bound on distillable entanglement*, Phys. Rev. A **60** (1999), 179–184.
- [171] M. Reed, B. Simon, *Methods of Modern Mathematical Physics, Vol. 1. Functional Analysis*, Academic Press, London, 1980.
- [172] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [173] M. B. Ruskai, *Inequalities for quantum entropy: A review with conditions for equality*, J. Math. Phys. **43** (2002), pp. 4358–4375.
- [174] M. B. Ruskai, S. Szarek, E. Werner, *A characterization of completely-positive trace-preserving maps on \mathcal{M}_2* , arXiv:quant-ph/0005004.
- [175] T. A. Sarymsakov, *Introduction to Quantum Probability Theory*, FAN, Tashkent, 1985. (in Russian)
- [176] M. Sasaki, S. M. Barnett, R. Jozsa, M. Osaki, O. Hirota, *Accessible information and optimal strategies for real symmetric quantum sources*, Phys. Rev. A **59** (1999), 3325; e-print quant-ph/9812062.
- [177] B. Schumacher, M. D. Westmoreland, *Sending classical information via noisy quantum channel*, Phys. Rev. A. **56** (1997), pp. 131–138.
- [178] B. Schumacher, M. D. Westmoreland, *Quantum privacy and quantum coherence*, Phys. Rev. Lett. **80** (1998), pp. 5695–5697; arXiv:quant-ph/9709058.
- [179] B. Schumacher, M. D. Westmoreland, *Optimal signal ensembles*, arXiv:quant-ph/9912122 (1999).
- [180] B. Schumacher, M. D. Westmoreland, *Approximate quantum error correction*, arXiv:quant-ph/0112106 (2001).
- [181] I. Segal, *Mathematical Problems of Relativistic Physics*, AMS, Providence, RI, 1963.
- [182] C. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Univ. Illinois Press, Urbana Ill., 1949.
- [183] M. E. Shirokov, *Entropic characteristics of subsets of states*, Izvestiya Math. **70** (2006), pp. 1265–1292; **71** (2007), pp. 181–218.
- [184] M. E. Shirokov, *On entropic quantities related to the classical capacity of infinite dimensional quantum channels*, Theory Probab. and Appl. **52** (2007), pp. 250–276.
- [185] M. E. Shirokov, *The Holevo capacity of infinite dimensional channels and the additivity problem*, Commun. Math. Phys. **262** (2006), pp. 137–159; arXiv:quant-ph/0408009.
- [186] M. E. Shirokov, *Entropy reduction of quantum measurements*, J. Math. Phys. **52** (2011), 052202; arXiv:1011.3127.

- [187] M. E. Shirokov, *A criterion for coincidence of the entanglement-assisted classical capacity and the Holevo capacity of a quantum channel*, Probl. Inform. Transmission. **48** (2012), pp. 3–19; arXiv:1202.3449v2 [quant-ph] (2012).
- [188] P. W. Shor, *Introduction to quantum algorithms*, arXiv:quant-ph/0005003 (2000).
- [189] P. W. Shor, *On the number of elements needed in a POVM attaining the accessible information*, arXiv:quant-ph/0009077 (2000).
- [190] P. W. Shor, *Additivity of the classical capacity of entanglement-breaking quantum channels*, J. Math. Phys., 2002, vol. 43, pp. 4334–4340; e-print quant-ph/0201149.
- [191] P. W. Shor, *The quantum channel capacity and coherent information*, Lecture Notes, MSRI Workshop on Quantum Computation, Berkeley, 2002.
- [192] P. W. Shor, *The adaptive classical capacity of a quantum channel, or information capacity of 3 symmetric pure states in three dimensions*, arXiv:quant-ph/0206058 (2002).
- [193] P. W. Shor, *Equivalence of additivity questions in quantum information theory*, Commun. Math. Phys. **246** (2004), pp. 453–472; arXiv:quant-ph/0305035.
- [194] P. W. Shor, *The classical capacity achievable by a quantum channel assisted by limited entanglement*, in: Quantum Information, Statistics, Probability, Ed. by O. Hirota, Rinton Press, Inc., Princeton, New Jersey 2004; quant-ph/0402129.
- [195] R. Simon, M. Salvadoray, G. S. Agarwal, *Gaussian states for finite number of Bosonic degrees of freedom*, Preprint, 1998.
- [196] G. Smith, *The private classical capacity with a symmetric side channel and its applications to quantum cryptography*, arXiv:0705.3838 [quant-ph] (2007).
- [197] G. Smith, J. A. Smolin, *Degenerate quantum codes for Pauli channels*, Phys. Rev. Lett. **98** (2007), 030501.
- [198] G. Smith, J. A. Smolin, *Detecting incapacity*, arXiv:1108.1807 [quant-ph] (2011).
- [199] G. Smith, J. A. Smolin, J. Yard, *Gaussian bosonic synergy: quantum communication via realistic channels of zero quantum capacity*, arXiv:1102.4580 [quant-ph] (2011).
- [200] G. Smith, J. Yard, *Quantum communication with zero-capacity channels*, Science **321** (2010), pp. 1812.
- [201] A. Steane: *Quantum computing*, Rept. Prog. Phys. **61** (1997), pp. 117–173.
- [202] W. F. Stinespring, *Positive functions on C^* -algebras*, Proc. Amer. Math. Soc. **6** (1955), pp. 211–316.
- [203] R. L. Stratonovich, *The amount of information transmitted by a quantum channel*, I, II, Izv. VUZ. Radiophysics **8** (1965), pp. 116–141. (in Russian)
- [204] R. L. Stratonovich, *The quantum generalization of optimal statistical estimation and testing hypothesis*, J. Stoch. **1** (1973), pp. 87–126.
- [205] R. L. Stratonovich, A. G. Vantsjan, *On asymptotically errorless decoding in pure quantum channels*, Probl. Control Inform. Theory, **7** (1978), pp. 161–174.
- [206] E. C. G. Sudarshan, P. M. Mathews, J. Rau, *Stochastic dynamics of quantum-mechanical systems*, Phys. Rev. **121** (1961), pp. 920–924.

- [207] J. A. Tropp, *User-friendly tail bounds for sums of random matrices*, arXiv:1104.4389 [math.PR] (2011).
- [208] A. Uhlmann, *The “transition probability” in the state space of a \ast -algebra*, Rep. Math. Phys. **9** (1976), pp. 273–279.
- [209] A. Uhlmann, *Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory*, Commun. Math. Phys. **54** (1977), pp. 21–32.
- [210] K. A. Valiev, *Quantum computers and quantum computations*, Physics-Uspekhi **48** (2005), pp. 1–36.
- [211] S. Verdú, *Fifty years of Shannon theory*, IEEE Trans. Inform. Theory **44** (1998), pp. 2057–2078.
- [212] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, NJ, 1955, (1932).
- [213] F. Verstraete, H. Verschelde, *On quantum channels*, arXiv:quant-ph/0202124 (2002).
- [214] C. Weedbrook, S. Pirandola, R. Garcia-Patron, N. J. Cerf, T. C. Ralph, J. H. Shapiro, S. Lloyd, *Gaussian quantum information*, arXiv:1110.3234 [quant-ph] (2011).
- [215] A. Wehrl, *General properties of entropy*, Rev. Mod. Phys. **50** (1978), pp. 221–260.
- [216] R. A. Werner, A. S. Holevo, *Counterexample to an additivity conjecture for output purity of quantum channels*, J. Math. Phys. **43** (2002), pp. 4353–4357.
- [217] R. F. Werner, M. M. Wolf, *Bound entangled Gaussian states*, arXiv:quant-ph/0009118.
- [218] H. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, New York, 1931, (1928).
- [219] A. S. Wightman, *Hilbert’s sixth problem: mathematical treatment of the axioms of physics*, Proc. Symp. in pure math. **28** (1977), pt. 1, pp. 147–240.
- [220] E. P. Wigner, *Group Theory*, Academic Press, New York, 1959.
- [221] M. M. Wilde, *From classical to quantum Shannon theory*, arXiv:1106.1445 [quant-ph] (2011).
- [222] J. Williamson, *On the algebraic problem concerning the normal forms of linear dynamical systems*, Am. J. Math **58** (1936), pp. 141; **59** (1937), pp. 599; **61** (1939), pp. 897.
- [223] A. Winter, *Coding theorem and strong converse for quantum channels*, IEEE Trans. Inform. Theory, **45** (1999), pp. 2481–2485.
- [224] A. Winter, *Compression of sources of probability distributions and density operators*, arXiv:quant-ph/0208131 (2002).
- [225] A. Winter, *The maximum output p -norm of quantum channels is not multiplicative for any $p > 2$* , arXiv:0707.0402 [quant-ph] (2007).
- [226] M. Wolf, G. Giedke, J. I. Cirac, *Extremality of Gaussian quantum states*, Phys. Rev. Lett. **96** (2006), 080502; arXiv:quant-ph/0509154.
- [227] M. M. Wolf, G. Giedke, O. Krueger, R. F. Werner, J. I. Cirac, *Gaussian entanglement of formation*, Phys. Rev. A **69** (2004), 052320; arXiv:quant-ph/0306177.

- [228] M. Wolf, D. Pérez-Garsía, G. Giedke, *Quantum capacities of Bosonic channels*, Phys. Rev. Lett. **98** (2007), 130501; quant-ph/0606132.
- [229] S. L. Woronowicz, *On the purification map*, Commun. Math. Phys. **30** (1973), pp. 55–67.
- [230] H. P. H. Yuen, M. Lax, *Multiple-parameter quantum estimation and measurement of non-selfadjoint observables*, IEEE Trans. Inform. Theory **19** (1973), pp. 740–750.
- [231] C. Zu, Y.-X. Wang, X.-Y. Chang, Z.-H. Wei, S.-Y. Zhang, L.-M. Duan, *Experimental demonstration of quantum gain in a zero-sum game*, New J. Phys. **14** (2012), 033002.

Index

- accessible information 78
- achievable rate 65
 - quantum information 210
 - quantum wiretap channel 221
- affine map 8
- asymptotic equipartition property 58
 - quantum 89
- Bell basis 50
- Bell state 45
- Bloch ball 15
- Born–von Neumann statistical
 - formula 18
- bound entanglement 53
- Bures distance 208
- canonical commutation relations (CCR) 276
 - discrete 125
 - Heisenberg 269, 277
 - Weyl–Segal 277
- canonical observables 277
- capacity
 - χ -capacity 157
 - classical 155
 - classical of c-q channel 75
 - constrained classical 249
 - entanglement-assisted classical 180
 - of classical channel 65
 - private classical 221
 - quantum 211
 - Shannon 62
- Cauchy–Schwarz inequality
 - noncommutative 22
- channel 113, 254
 - anti-degradable 218
 - bistochastic 113
 - classical 60
 - classical wiretap 69
 - classical-quantum (c-q) 74, 114, 249
 - complementary 120
 - completely depolarizing 116
 - covariant 124
 - degradable 218
 - dephasing 123
 - depolarizing 124
 - entanglement-binding 116
 - entanglement-breaking 115, 260
 - Gaussian 297
 - linear bosonic 297
 - orthogonal convex sum of 123
 - perfectly reversible 201
 - PPT 116
 - pure loss 324
 - quantum erasure 123
 - quantum wiretap 220
 - quantum-classical (q-c) 114, 184
 - reverse 197, 201
 - transpose-depolarizing 171
 - trine 77
 - weakly complementary 120
- characteristic function 284
 - of observable 301
 - of state 284
- Choi–Jamiołkowski correspondence 114
- Clauser–Horne–Shimony–Holt
 - inequality 43
- coarse-graining 12
- code 74, 249
 - classical 64
 - classical capacity 155
 - error correcting 197
- commutation matrix 21
- commutator 18
- compatible observables 12
 - joint probability distribution of 19
- complementarity 19
- completely positive map 106
 - rank 109
- concave function 8

- conditionally typical
 - projector 91
 - word 67
- convex
 - combination 8
 - function 8
 - set 8
- covariance matrix 21, 285
- data compression
 - classical 58
 - quantum 89
- data-processing inequalities
 - classical 62
 - quantum 149
- decision rule
 - classical 64
 - quantum 74
- density operator 244
- ensemble 156
 - average state of 256
 - generalized 256
 - of quantum states 143
 - statistical (sample) 11
- ensemble-observable duality 185
- entanglement of formation 143
- entropy
 - binary 62
 - classical 57
 - classical conditional 60
 - classical relative 60
 - quantum conditional 146
 - quantum Rényi 171
 - quantum relative 132
 - von Neumann 75
- entropy exchange 147
- entropy gain 137
 - minimal 164
- expectation 18
- Fano's inequality 66
- fidelity 204, 208
 - average 89
- gain of entanglement assistance 182
- gauge
 - covariant channel 308
 - group 283
 - invariance 283
- generalized H-theorem 136
- Hadamard gate 51
- Heisenberg picture 104
- information
 - coherent 200
 - quantum mutual 149
 - Shannon 60
- Jordan product 21
- Kadison inequality 107
- Kraus representation 110
 - minimal 110
- Lindblad–Ozawa inequality 166
- locality (separability) 43
- maximal likelihood 27
- mean vector 285
- measurement
 - ideal 117
 - indirect 118
- mixture
 - of observables 22
 - of states 12, 14
- Naimark's dilation theorem 36
- no cloning 50
- norm
 - operator 7
 - trace 6
- observable 11, 260
 - classical
 - sharp, unsharp 10
 - Gaussian 302
 - probability distribution of 11
 - quantum 17
 - probability distribution of 17
 - real 18
 - sharp 17, 261

- operator 272
 - adjoint 4
 - antiunitary 104
 - bounded 243
 - creation-annihilation 272, 281
 - energy 272
 - essentially self-adjoint 268
 - Hermitian 4
 - isometric 4
 - normal 5
 - number of quanta 272, 283
 - of complex structure 279
 - of the type \mathfrak{F} 246
 - positive 5
 - self-adjoint 268
 - trace-class 243
 - unitary 4
- operator-convex function 151
- operator-monotone function 151
- overcomplete system 23

- Pauli matrices 15
- polar decomposition 6
- polarization identity 4
- positive partial transpose (PPT) 40
- pre-inner product 36
- probability operator-valued measure (POVM) 17, 260
 - orthogonal 261
- projector 4
- purification 39

- qubit 15

- reference system 113
- reliability function 95
- resolution of the identity 17
 - orthogonal 5, 18

- Schmidt decomposition 38
- Schrödinger picture 104
- Shor code 197
- simplex 8
- spectral measure 261
- spectral theorem 5, 268
- state 11
 - chaotic 15
 - coherent 272, 288
 - Gaussian 285
 - centered 286
 - elementary 285
 - maximally entangled 38
 - partial 35
 - posterior 48, 117
 - pure
 - classical 10
 - entangled 38
 - quantum 14
 - separable (unentangled) 40, 259
 - singlet 42
 - thermal 273
 - vacuum 272, 288
- statistical model 12
 - Kolmogorov 13
 - Wald 13
- Stinespring representation 107
 - minimal 109
- stochastically compatible observables 13
- Stone's Theorem 268
- strongly typical
 - projector 192
 - word 192
- support of operator 6
- symplectic 278
 - basis 279
 - form 278
 - space 279
 - transformation 279

- tensor product 34
- test 23
- trace 6
 - partial 35
- typical
 - projector 88
 - subspace 88
 - word 58

- uncertainty relation
 - Heisenberg 270
 - Schrödinger–Robertson 21

- variance 21
- von Neumann–Lüders projection
 - postulate 117

weak convergence
 of operators 244
 of probability measures 252

Weyl operators 277
 discrete 125
Wigner's Theorem 104